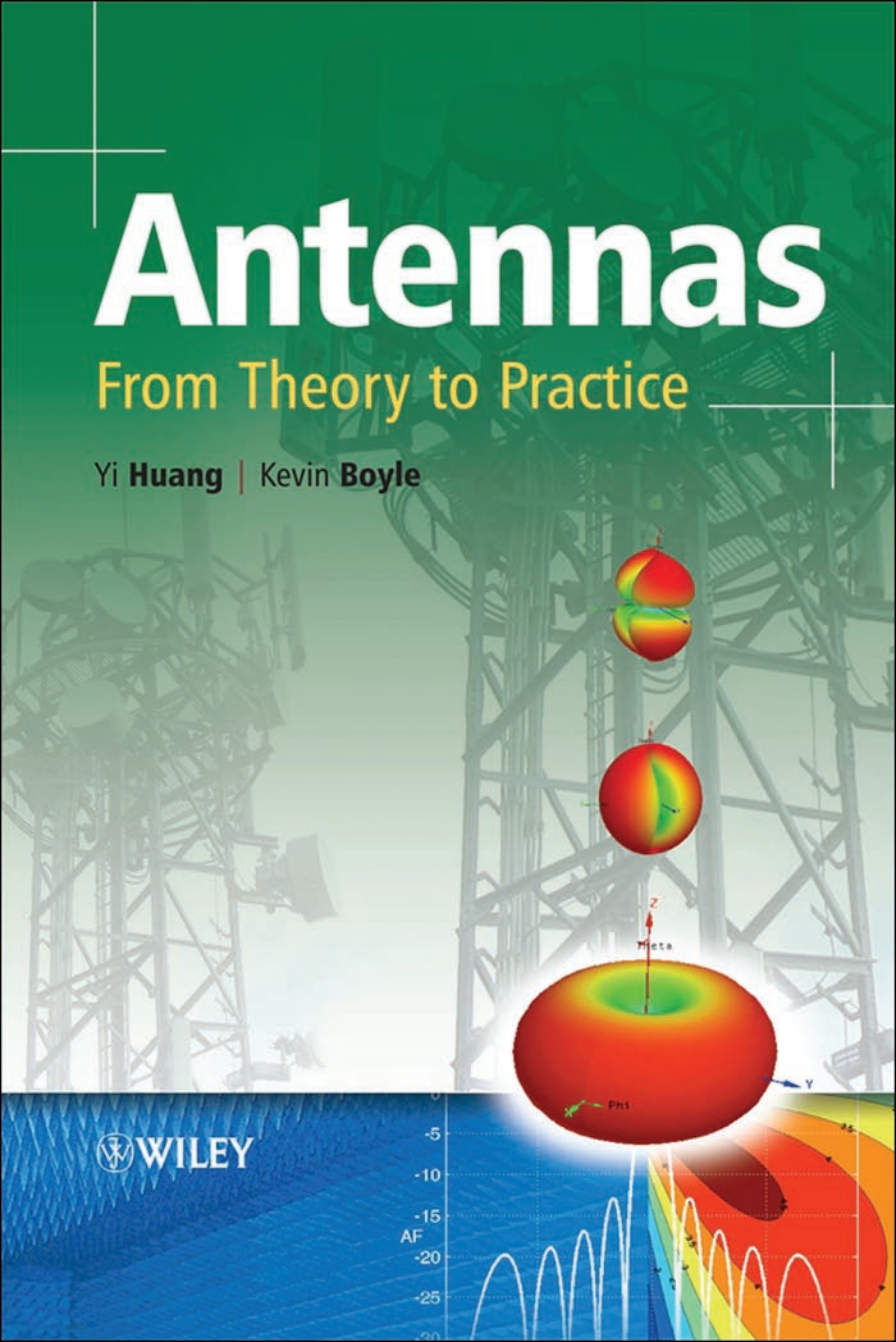
Antennas

From Theory to Practice

Yi **Huang** | Kevin **Boyle**



WILEY

VISIT…

ANTENNAS
FROM THEORY TO PRACTICE

Yi Huang
University of Liverpool, UK

Kevin Boyle
NXP Semiconductors, UK

WILEY
A John Wiley and Sons, Ltd, Publication

ANTENNAS

ANTENNAS
FROM THEORY TO PRACTICE

Yi Huang
University of Liverpool, UK

Kevin Boyle
NXP Semiconductors, UK

WILEY
A John Wiley and Sons, Ltd, Publication

This edition first published 2008
© 2008 John Wiley & Sons Ltd

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Huang, Yi.
Antennas : from theory to practice / Yi Huang, Kevin Boyle.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-51028-5 (cloth)
1. Antennas (Electronics) I. Boyle, Kevin. II. Title.
TK7871.6.H79 2008
621.382′4—dc22 2008013164

A catalogue record for this book is available from the British Library.

ISBN 978-0-470-51028-5 (HB)

Typeset in 10/12pt Times by Aptara Inc., New Delhi, India.
Printed in Singapore by Markono Print Media Pte Ltd, Singapore.

Contents

Preface **xi**

Acronyms and Constants **xiii**

1 Introduction **1**
1.1 A Short History of Antennas 1
1.2 Radio Systems and Antennas 4
1.3 Necessary Mathematics 6
1.3.1 Complex Numbers 6
1.3.2 Vectors and Vector Operation 7
1.3.3 Coordinates 10
1.4 Basics of Electromagnetics 11
1.4.1 The Electric Field 12
1.4.2 The Magnetic Field 15
1.4.3 Maxwell's Equations 16
1.4.4 Boundary Conditions 19
1.5 Summary 21
References 21
Problems 21

2 Circuit Concepts and Transmission Lines **23**
2.1 Circuit Concepts 23
2.1.1 Lumped and Distributed Element Systems 25
2.2 Transmission Line Theory 25
2.2.1 Transmission Line Model 25
2.2.2 Solutions and Analysis 28
2.2.3 Terminated Transmission Line 32
2.3 The Smith Chart and Impedance Matching 41
2.3.1 The Smith Chart 41
2.3.2 Impedance Matching 44
2.3.3 The Quality Factor and Bandwidth 51
2.4 Various Transmission Lines 55
2.4.1 Two-wire Transmission Line 56
2.4.2 Coaxial Cable 57
2.4.3 Microstrip Line 60

2.4.4 Stripline 63
2.4.5 Coplanar Waveguide (CPW) 66
2.4.6 Waveguide 68
2.5 Connectors 70
2.6 Summary 74
References 74
Problems 74

3 Field Concepts and Radio Waves **77**
3.1 Wave Equation and Solutions 77
3.1.1 Discussion on Wave Solutions 79
3.2 The Plane Wave, Intrinsic Impedance and Polarization 80
3.2.1 The Plane Wave and Intrinsic Impedance 80
3.2.2 Polarization 82
3.3 Radio Wave Propagation Mechanisms 83
3.3.1 Reflection and Transmission 83
3.3.2 Diffraction and Huygens's Principle 91
3.3.3 Scattering 92
3.4 Radio Wave Propagation Characteristics in Media 93
3.4.1 Media Classification and Attenuation 93
3.5 Radio Wave Propagation Models 97
3.5.1 Free Space Model 97
3.5.2 Two-ray Model/Plane Earth Model 98
3.5.3 Multipath Models 99
3.6 Comparison of Circuit Concepts and Field Concepts 101
3.6.1 Skin Depth 101
3.7 Summary 104
References 104
Problems 104

4 Antenna Basics **107**
4.1 Antennas to Radio Waves 107
4.1.1 Near Field and Far Field 108
4.1.2 Antenna Parameters from the Field Point of View 112
4.2 Antennas to Transmission Lines 122
4.2.1 Antenna Parameters from the Circuit Point of View 122
4.3 Summary 125
References 126
Problems 126

5 Popular Antennas **129**
5.1 Wire-Type Antennas 129
5.1.1 Dipoles 129
5.1.2 Monopoles and Image Theory 137
5.1.3 Loops and the Duality Principle 141
5.1.4 Helical Antennas 147

5.1.5 Yagi–Uda Antennas 152
5.1.6 Log-Periodic Antennas and Frequency-Independent Antennas 157
5.2 Aperture-Type Antennas 163
5.2.1 Fourier Transforms and the Radiated Field 163
5.2.2 Horn Antennas 169
5.2.3 Reflector and Lens Antennas 175
5.2.4 Slot Antennas and Babinet's Principle 180
5.2.5 Microstrip Antennas 184
5.3 Antenna Arrays 191
5.3.1 Basic Concept 192
5.3.2 Isotropic Linear Arrays 193
5.3.3 Pattern Multiplication Principle 199
5.3.4 Element Mutual Coupling 200
5.4 Some Practical Considerations 203
5.4.1 Transmitting and Receiving Antennas: Reciprocity 203
5.4.2 Baluns and Impedance Matching 205
5.4.3 Antenna Polarization 206
5.4.4 Radomes, Housings and Supporting Structures 208
5.5 Summary 211
References 211
Problems 212

6 Computer-Aided Antenna Design and Analysis 215
6.1 Introduction 215
6.2 Computational Electromagnetics for Antennas 217
6.2.1 Method of Moments (MoM) 218
6.2.2 Finite Element Method (FEM) 228
6.2.3 Finite-Difference Time Domain (FDTD) Method 229
6.2.4 Transmission Line Modeling (TLM) Method 230
6.2.5 Comparison of Numerical Methods 230
6.2.6 High-Frequency Methods 232
6.3 Examples of Computer-Aided Design and Analysis 233
6.3.1 Wire-type Antenna Design and Analysis 233
6.3.2 General Antenna Design and Analysis 243
6.4 Summary 251
References 251
Problems 252

7 Antenna Manufacturing and Measurements 253
7.1 Antenna Manufacturing 253
7.1.1 Conducting Materials 253
7.1.2 Dielectric Materials 255
7.1.3 New Materials for Antennas 255
7.2 Antenna Measurement Basics 256
7.2.1 Scattering Parameters 256
7.2.2 Network Analyzers 258

7.3 Impedance, S_{11}, VSWR and Return Loss Measurement 261
7.3.1 Can I Measure These Parameters in My Office? 261
7.3.2 Effects of a Small Section of a Transmission Line or a Connector 262
7.3.3 Effects of Packages on Antennas 262
7.4 Radiation Pattern Measurements 263
7.4.1 Far-Field Condition 264
7.4.2 Open-Area Test Sites (OATS) 265
7.4.3 Anechoic Chambers 267
7.4.4 Compact Antenna Test Ranges (CATR) 268
7.4.5 Planar and Cylindrical Near-Field Chambers 270
7.4.6 Spherical Near-Field Chambers 270
7.5 Gain Measurements 272
7.5.1 Comparison with a Standard-Gain Horn 272
7.5.2 Two-Antenna Measurement 272
7.5.3 Three-Antenna Measurement 273
7.6 Miscellaneous Topics 273
7.6.1 Efficiency Measurements 273
7.6.2 Reverberation Chambers 274
7.6.3 Impedance De-embedding Techniques 275
7.6.4 Probe Array in Near-Field Systems 276
7.7 Summary 281
References 281
Problems 282

8 Special Topics 283
8.1 Electrically Small Antennas 283
8.1.1 The Basics and Impedance Bandwidth 283
8.1.2 Antenna Size-Reduction Techniques 299
8.2 Mobile Antennas, Antenna Diversity and Human Body Effects 304
8.2.1 Introduction 304
8.2.2 Mobile Antennas 305
8.2.3 Antenna Diversity 318
8.2.4 User Interaction 325
8.3 Multiband and Ultra-Wideband Antennas 334
8.3.1 Introduction 334
8.3.2 Multiband Antennas 334
8.3.3 Wideband Antennas 337
8.4 RFID Antennas 340
8.4.1 Introduction 340
8.4.2 Near-Field Systems 343
8.4.3 Far-Field Systems 349
8.5 Reconfigurable Antennas 352
8.5.1 Introduction 352
8.5.2 Switching and Variable-Component Technologies 352
8.5.3 Resonant Mode Switching/Tuning 354

8.5.4 Feed Network Switching/Tuning 355
8.5.5 Mechanical Reconfiguration 355
8.6 Summary 356
References 356

Index **357**

Preface

As an essential element of a radio system, the antenna has always been an interesting but difficult subject for radio frequency (RF) engineering students and engineers. Many good books on antennas have been published over the years and some of them were used as our major references.

This book is different from other antenna books. It is especially designed for people who know little about antennas but would like to learn this subject from the very basics to practical antenna analysis, design and measurement within a relatively short period of time. In order to gain a comprehensive understanding of antennas, one must know about transmission lines and radio propagation. At the moment, people often have to read a number of different books, which may not be well correlated. Thus, it is not the most efficient way to study the subject. In this book we put all the necessary information about antennas into a single volume and try to examine antennas from both the circuit point of view and the field point of view. The book covers the basic transmission line and radio propagation theories, which are then used to gain a good understanding of antenna basics and theory. Various antennas are examined and design examples are presented. Particular attention is given to modern computer-aided antenna design. Both basic and advanced computer software packages are used in examples to illustrate how they can be used for antenna analysis and design. Antenna measurement theory and techniques are also addressed. Some special topics on the latest antenna development are covered in the final chapter.

The material covered in the book is mainly based on a successful short course on antennas for practising professionals at the University of Oxford and the Antennas module for students at the University of Liverpool. The book covers important and timely issues involving modern practical antenna design and theory. Many examples and questions are given in each chapter. It is an ideal textbook for university antenna courses, professional training courses and self-study. It is also a valuable reference for engineers and designers who work with RF engineering, radar and radio communications.

The book is organized as follows:

Chapter 1: Introduction. The objective of this chapter is to introduce the concept of antennas and review essential mathematics and electromagnetics, especially Maxwell's equations. Material properties (permittivity, permeability and conductivity) are discussed and some common ones are tabulated.

Chapter 2: Circuit Concepts and Transmission Lines. The concepts of lumped and distributed systems are established. The focus is placed on the fundamentals and characteristics of transmission lines. A comparison of various transmission lines and connectors is presented. The Smith Chart, impedance matching and bandwidth are also addressed in this chapter.

Chapter 3: Field Concepts and Radio Waves. Field concepts, including the plane wave, intrinsic impedance and polarization, are introduced and followed by a discussion on radio propagation mechanisms and radio wave propagation characteristics in various media. Some basic radio propagation models are introduced, and circuit concepts and field concepts are compared at the end of this chapter.

Chapter 4: Antenna Basics. The essential and important parameters of an antenna (such as the radiation pattern, gain and input impedance) are addressed from both the circuit point of view and field point of view. Through this chapter, you will become familiar with antenna language, understand how antennas work and know what design considerations are.

Chapter 5: Popular Antennas. In this long chapter, some of the most popular antennas (wire-type, aperture-type and array antennas) are examined and analyzed using relevant antenna theories. The aim is to see why they have become popular, what their major features and properties are (including advantages and disadvantages) and how they should be designed.

Chapter 6: Computer-Aided Antenna Design and Analysis. The aim of this special and unique chapter is to give a brief review of antenna-modeling methods and software development, introduce the basic theory behind computer simulation tools and demonstrate how to use industry standard software to analyze and design antennas. Two software packages (one is simple and free) are presented with step-by-step illustrations.

Chapter 7: Antenna Manufacturing and Measurements. This is another practical chapter to address two important issues: how to make an antenna and how to conduct antenna measurement, with a focus placed on the measurement. It introduces S-parameters and equipment. A good overview of the possible measurement systems is provided with an in-depth example. Some measurement techniques and problems are also presented.

Chapter 8: Special Topics. This final chapter presents some of the latest important developments in antennas. It covers mobile antennas and antenna diversity, RFID antennas, multiband and broadband antennas, reconfigurable antennas and electrically small antennas. Both the theory and practical examples are given.

The authors are indebted to the many individuals who provided useful comments, suggestions and assistance to make this book a reality. In particular, we would like to thank Shahzad Maqbool, Barry Cheeseman and Yang Lu at the University of Liverpool for constructive feedback and producing figures, Staff at Wiley for their help and critical review of the book, Lars Foged at SATIMO and Mike Hillbun at Diamond Engineering for their contribution to Chapter 7 and the individuals and organizations who have provided us with their figures or allowed us to reproduce their figures.

Yi Huang and Kevin Boyle

Acronyms and Constants

ε_0	8.85419×10^{-12} F/m
μ_0	$4\pi \times 10^{-7}$ H/m
η_0	$\approx 377\ \Omega$

AC	Alternating current
AF	Antenna factor
AM	Amplitude modulation
AR	Axial ratio
AUT	Antenna under test
BER	Bit error rate
CAD	Computer-aided design
CATR	Compact antenna test range
CDF	Cumulative distribution function
CEM	Computational electromagnetics
CP	Circular polarization
CPW	Coplanar waveguide
DC	Direct current
DCS	Digital cellular system
DRA	Dielectric resonant antenna
DUT	Device under test
EGC	Equal gain combining
EIRP	Effective isotropic radiated power
EM	Electromagnetic
EMC	Electromagnetic compatibility
ERP	Effective radiated power
FDTD	Finite-difference time domain
FEM	Finite element method
FNBW	First null beamwidth
GPS	Global positioning system
GSM	Global System for Mobile communications
GTD	Geometrical theory of diffraction
HPBW	Half-power beamwidth
HW	Hansen–Woodyard (condition)
ISI	Inter-symbol interference

LCP	Left-hand circular polarization
	Liquid crystal polymer
LPDA	Log-periodic dipole antenna
MEMS	Micro electromechanical systems
MIMO	Multiple-in, multiple-out
MMIC	Monolithic microwave integrated circuits
MoM	Method of moments
MRC	Maximal ratio combining
NEC	Numerical electromagnetic code
OATS	Open area test site
PCB	Printed circuit board
PDF	Power density function
	Probability density function
PIFA	Planar inverted F antenna
PO	Physical optics
PTFE	Polytetrafluoroethylene
RAM	Radio-absorbing material
RCP	Right-hand circular polarization
RCS	Radar cross-section
RF	Radio frequency
RFID	Radio frequency identification
RMS	Root mean square
SAR	Specific absorption rate
SC	Selection combining
SI units	International system of units (metric system)
SLL	Side-lobe level
SNR	Signal-to-noise ratio
SWC	Switch combining
TE	Transverse electric (mode/field)
TEM	Transverse electromagnetic (mode/field)
TM	Transverse magnetic (mode/field)
TV	Television
UHF	Ultra-high frequency
UTD	Uniform theory of diffraction
UWB	Ultra-wide band
VHF	Very high frequency
VNA	Vector network analyzer
VSWR	Voltage standing wave ratio
WLAN	Wireless local area network
WiMax	Worldwide interoperability of microwave access

1

Introduction

1.1 A Short History of Antennas

Work on antennas started many years ago. The first well-known satisfactory antenna experiment was conducted by the German physicist Heinrich Rudolf Hertz (1857–1894), pictured in Figure 1.1. The SI (International Standard) frequency unit, the Hertz, is named after him. In 1887 he built a system, as shown in Figure 1.2, to produce and detect radio waves. The original intention of his experiment was to demonstrate the existence of electromagnetic radiation.

In the transmitter, a variable voltage source was connected to a dipole (a pair of one-meter wires) with two conducting balls (capacity spheres) at the ends. The gap between the balls could be adjusted for circuit resonance as well as for the generation of sparks. When the voltage was increased to a certain value, a spark or break-down discharge was produced. The receiver was a simple loop with two identical conducting balls. The gap between the balls was carefully tuned to receive the spark effectively. He placed the apparatus in a darkened box in order to see the spark clearly. In his experiment, when a spark was generated at the transmitter, he also observed a spark at the receiver gap at almost the same time. This proved that the information from location A (the transmitter) was transmitted to location B (the receiver) in a wireless manner – by electromagnetic waves.

The information in Hertz's experiment was actually in binary digital form, by tuning the spark on and off. This could be considered the very first digital wireless system, which consisted of two of the best-known antennas: the dipole and the loop. For this reason, the dipole antenna is also called the Hertz (dipole) antenna.

Whilst Heinrich Hertz conducted his experiments in a laboratory and did not quite know what radio waves might be used for in practice, Guglielmo Marconi (1874–1937, pictured in Figure 1.3), an Italian inventor, developed and commercialized wireless technology by introducing a radiotelegraph system, which served as the foundation for the establishment of numerous affiliated companies worldwide. His most famous experiment was the transatlantic transmission from Poldhu, UK to St Johns, Newfoundland in Canada in 1901, employing untuned systems. He shared the 1909 Nobel Prize for Physics with Karl Ferdinand Braun 'in recognition of their contributions to the development of wireless telegraphy'. Monopole antennas (near quarter-wavelength) were widely used in Marconi's experiments; thus vertical monopole antennas are also called Marconi antennas.

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

Figure 1.1 Heinrich Rudolf Hertz

During World War II, battles were won by the side that was first to spot enemy aeroplanes, ships or submarines. To give the Allies an edge, British and American scientists developed radar technology to 'see' targets from hundreds of miles away, even at night. The research resulted in the rapid development of high-frequency radar antennas, which were no longer just wire-type antennas. Some aperture-type antennas, such as reflector and horn antennas, were developed, an example is shown in Figure 1.4.

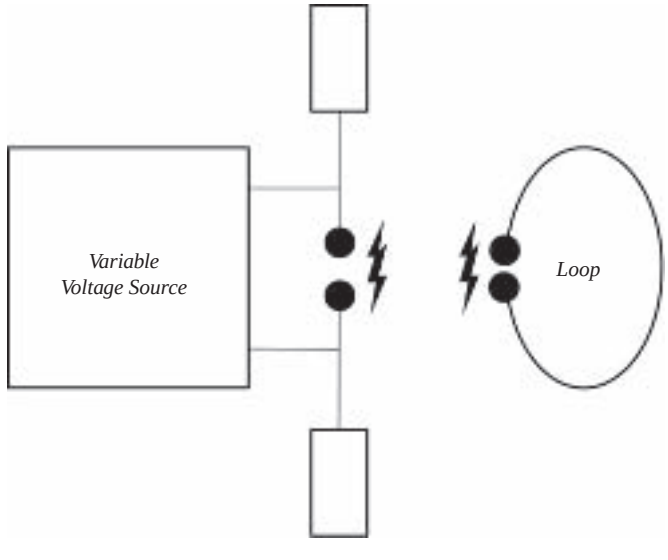


Figure 1.2 1887 experimental set-up of Hertz's apparatus

Figure 1.3 Guglielmo Marconi

Broadband, circularly polarized antennas, as well as many other types, were subsequently developed for various applications. Since an antenna is an essential device for any radio broadcasting, communication or radar system, there has always been a requirement for new and better antennas to suit existing and emerging applications.

More recently, one of the main challenges for antennas has been how to make them broadband and small enough in size for wireless mobile communications systems. For example, WiMAX (worldwide interoperability for microwave access) is one of the latest systems aimed at providing high-speed wireless data communications ($>$10 Mb/s) over long distances from point-to-point links to full mobile cellular-type access over a wide frequency band. The original WiMAX standard in IEEE 802.16 specified 10 to 66 GHz as the WiMAX band; IEEE 802.16a

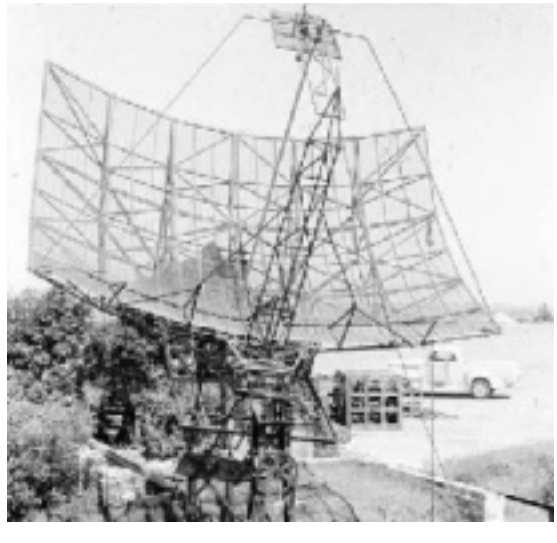


Figure 1.4 World War II radar (Reproduced by permission of CSIRO Australia Telescope National Facility)

was updated in 2004 to 802.16-2004 and added 2 to 11 GHz as an additional frequency range. The frequency bandwidth is extremely wide although the most likely frequency bands to be used initially will be around 3.5 GHz, 2.3/2.5 GHz and 5 GHz.

The UWB (ultra-wide band) wireless system is another example of recent broadband radio communication systems. The allocated frequency band is from 3.1 to 10.6 GHz. The beauty of the UWB system is that the spectrum, which is normally very expensive, can be used free of charge but the power spectrum density is limited to -41.3 dBm/MHz. Thus, it is only suitable for short-distance applications. The antenna design for these systems faces many challenging issues.

The role of antennas is becoming increasingly important. In some systems, the antenna is now no longer just a simple transmitting/receiving device, but a device which is integrated with other parts of the system to achieve better performance. For example, the MIMO (multiple-in, multiple-out) antenna system has recently been introduced as an effective means to combat multipath effects in the radio propagation channel and increase the channel capacity, where several coordinated antennas are required.

Things have been changing quickly in the wireless world. But one thing has never changed since the very first antenna was made: the antenna is a practical engineering subject. It will remain an engineering subject. Once an antenna is designed and made, it must be tested. How well it works is not just determined by the antenna itself, it also depends on the other parts of the system and the environment. The standalone antenna performance can be very different from that of an installed antenna. For example, when a mobile phone antenna is designed, we must take the case, other parts of the phone and even our hands into account to ensure that it will work well in the real world. The antenna is an essential device of a radio system, but not an isolated device! This makes it an interesting and challenging subject.

1.2 Radio Systems and Antennas

A radio system is generally considered to be an electronic system which employs radio waves, a type of electromagnetic wave up to GHz frequencies. An *antenna*, as an essential part of a radio system, is defined as a device which can radiate and receive electromagnetic energy in an efficient and desired manner. It is normally made of metal, but other materials may also be used. For example, ceramic materials have been employed to make dielectric resonator antennas (DRAs). There are many things in our lives, such as power leads, that can radiate and receive electromagnetic energy but they cannot be viewed as antennas because the electromagnetic energy is not transmitted or received in an efficient and desired manner, and because they are not a part of a radio system.

Since radio systems possess some unique and attractive advantages over wired systems, numerous radio systems have been developed. TV, radar and mobile radio communication systems are just some examples. The advantages include:

- **mobility:** this is essential for mobile communications;
- **good coverage:** the radiation from an antenna can cover a very large area, which is good for TV and radio broadcasting and mobile communications;
- **low pathloss:** this is frequency dependent. Since the loss of a transmission line is an exponential function of the distance (the loss in dB $=$ distance $\times$ per unit loss in dB) and the loss

of a radio wave is proportional to the distance squared (the loss in dB $= 20 \log_{10}$ (distance)), the pathloss of radio waves can be much smaller than that of a cable link. For example, assume that the loss is 10 dB for both a transmission line and a radio wave over 100 m; if the distance is now increased to 1000 m, the loss for the transmission line becomes $10 \times 10 =$ 100 dB but the loss for the radio link is just $10 + 20 = 30$ dB! This makes the radio link extremely attractive for long-distance communication. It should be pointed out that optical fibers are also employed for long-distance communications since they are of very low loss and ultra-wide bandwidth.

Figure 1.5 illustrates a typical radio communication system. The source information is normally modulated and amplified in the transmitter and then passed on to the transmit antenna via a transmission line, which has a typical characteristic impedance (explained in the next chapter) of 50 ohms. The antenna radiates the information in the form of an electromagnetic wave in an efficient and desired manner to the destination, where the information is picked up by the receive antenna and passed on to the receiver via another transmission line. The signal is demodulated and the original message is then recovered at the receiver.

Thus, the antenna is actually a transformer that transforms electrical signals (voltages and currents from a transmission line) into electromagnetic waves (electric and magnetic fields), or vice versa. For example, a satellite dish antenna receives the radio wave from a satellite and transforms it into electrical signals which are output to a cable to be further processed. Our eyes may be viewed as another example of antennas. In this case, the wave is not a radio wave but an optical wave, another form of electromagnetic wave which has much higher frequencies.

Now it is clear that the antenna is actually a transformer of voltage/current to electric/magnetic fields, it can also be considered a bridge to link the radio wave and transmission line. An *antenna system* is defined as the combination of the antenna and its feed line. As an antenna is usually connected to a transmission line, how to best make this connection is a subject of interest, since the signal from the feed line should be radiated into the space in an efficient and desired way. Transmission lines and radio waves are, in fact, two different subjects in engineering. To understand antenna theory, one has to understand transmission lines and radio waves, which will be discussed in detail in Chapters 2 and 3 respectively.

In some applications where space is very limited (such as hand-portables and aircraft), it is desirable to integrate the antenna and its feed line. In other applications (such as the reception of TV broadcasting), the antenna is far away from the receiver and a long transmission line has to be used.

Unlike other devices in a radio system (such as filters and amplifiers), the antenna is a very special device; it deals with electrical signals (voltages and currents) as well as electromagnetic waves (electric fields and magnetic fields), making antenna design an interesting and difficult

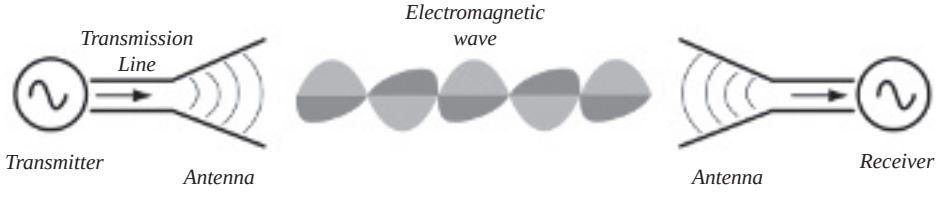


Figure 1.5 A typical radio system

subject. For different applications, the requirements on the antenna may be very different, even for the same frequency band.

In conclusion, the subject of antennas is about how to design a suitable device which will be well matched with its feed line and radiate/receive the radio waves in an efficient and desired manner.

1.3 Necessary Mathematics

To understand antenna theory thoroughly requires a considerable amount of mathematics. However, the intention of this book is to provide the reader with a solid foundation in antenna theory and apply the theory to practical antenna design. Here we are just going to introduce and review the essential and important mathematics required for this book. More in-depth study materials can be obtained from other references [1, 2].

1.3.1 Complex Numbers

In mathematics, a complex number, Z, consists of real and imaginary parts, that is

$$Z = R + jX \tag{1.1}$$

where R is called the real part of the complex number Z, i.e. Re(Z), and X is defined as the imaginary part of Z, i.e. Im(Z). Both R and X are real numbers and j (not the traditional notation i in mathematics to avoid confusion with a changing current in electrical engineering) is the imaginary unit and is defined by

$$j = \sqrt{-1} \tag{1.2}$$

Thus

$$j^2 = -1 \tag{1.3}$$

Geometrically, a complex number can be presented in a two-dimensional plane where the imaginary part is found on the vertical axis whilst the real part is presented by the horizontal axis, as shown in Figure 1.6.

In this model, multiplication by -1 corresponds to a rotation of 180 degrees about the origin. Multiplication by j corresponds to a 90-degree rotation anti-clockwise, and the equation $j^2 = -1$ is interpreted as saying that if we apply two 90-degree rotations about the origin, the net result is a single 180-degree rotation. Note that a 90-degree rotation clockwise also satisfies this interpretation.

Another representation of a complex number Z uses the amplitude and phase form:

$$Z = Ae^{j\varphi} \tag{1.4}$$

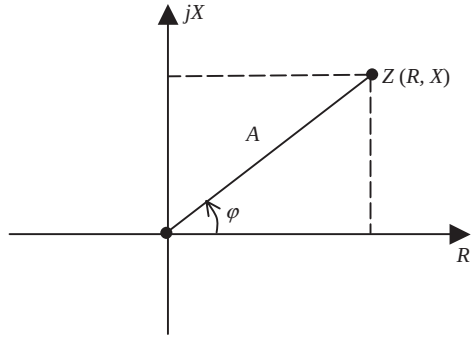


Figure 1.6 The complex plane

where A is the amplitude and φ is the phase of the complex number Z; these are also shown in Figure 1.6. The two different representations are linked by the following equations:

$$\begin{aligned} &Z = R + jX = Ae^{j\varphi}; \\ &A = \sqrt{R^2 + X^2}, \quad \varphi = \tan^{-1}(X/R) \\ &R = A\cos\varphi, \quad X = A\sin\varphi \end{aligned} \tag{1.5}$$

1.3.2 Vectors and Vector Operation

A scalar is a one-dimensional quantity which has magnitude only, whereas a complex number is a two-dimensional quantity. A vector can be viewed as a three-dimensional (3D) quantity, and a special one – it has both a magnitude and a direction. For example, force and velocity are vectors (whereas speed is a scalar). A position in space is a 3D quantity, but it does not have a direction, thus it is not a vector. Figure 1.7 is an illustration of vector A in Cartesian

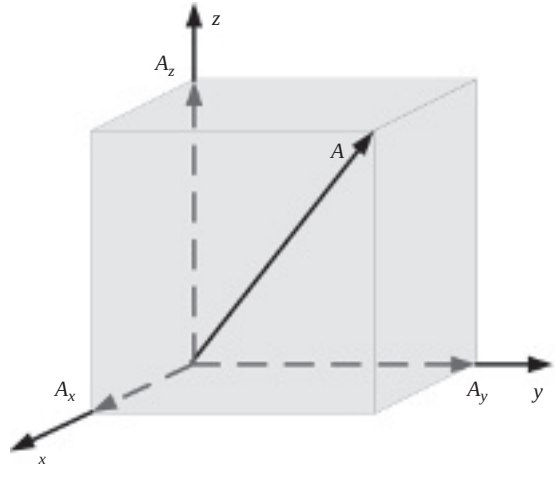


Figure 1.7 Vector ***A*** in Cartesian coordinates

coordinates. It has three orthogonal components (A_x, A_y, A_z) along the x, y and z directions, respectively. To distinguish vectors from scalars, the letter representing the vector is printed in bold, for example $\boldsymbol{A}$ or $\boldsymbol{a}$, and a unit vector is printed in bold with a hat over the letter, for example $\hat{\boldsymbol{x}}$ or $\hat{\boldsymbol{n}}$.

The magnitude of vector $\boldsymbol{A}$ is given by

$$|\boldsymbol{A}| = A = \sqrt{A_x^2 + A_y^2 + A_z^2} \tag{1.6}$$

Now let us consider two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\boldsymbol{A} = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}} + A_z\hat{\boldsymbol{z}}$$
$$\boldsymbol{B} = B_x\hat{\boldsymbol{x}} + B_y\hat{\boldsymbol{y}} + B_z\hat{\boldsymbol{z}}$$

The addition and subtraction of vectors can be expressed as

$$\begin{aligned} \boldsymbol{A} + \boldsymbol{B} &= (A_x + B_x)\hat{\boldsymbol{x}} + (A_y + B_y)\hat{\boldsymbol{y}} + (A_z + B_z)\hat{\boldsymbol{z}} \\ \boldsymbol{A} - \boldsymbol{B} &= (A_x - B_x)\hat{\boldsymbol{x}} + (A_y - B_y)\hat{\boldsymbol{y}} + (A_z - B_z)\hat{\boldsymbol{z}} \end{aligned} \tag{1.7}$$

Obviously, the addition obeys the *commutative law*, that is $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$.

Figure 1.8 shows what the addition and subtraction mean geometrically. A vector may be multiplied or divided by a scalar. The magnitude changes but its direction remains the same. However, the multiplication of two vectors is complicated. There are two types of multiplication: the dot product and the cross product.

The *dot product* of two vectors is defined as

$$\boldsymbol{A} \bullet \boldsymbol{B} = |\boldsymbol{A}|\,|\boldsymbol{B}| \cos\theta = A_x B_x + A_y B_y + A_z B_z \tag{1.8}$$

where θ is the angle between vector $\boldsymbol{A}$ and vector $\boldsymbol{B}$ and $\cos\theta$ is also called the direction cosine. The dot • between $\boldsymbol{A}$ and $\boldsymbol{B}$ indicates the dot product, which results in a scalar; thus, it is also called a *scalar product*. If the angle θ is zero, $\boldsymbol{A}$ and $\boldsymbol{B}$ are in parallel – the dot product is

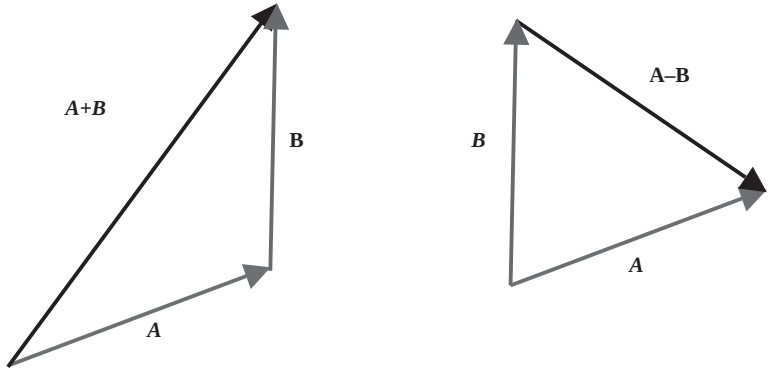


Figure 1.8 Vector addition and subtraction

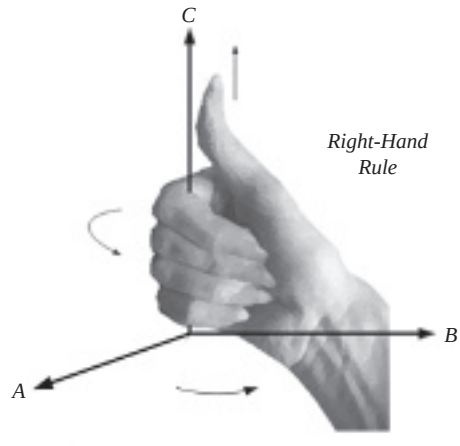


Figure 1.9 The cross product of vectors $\boldsymbol{A}$ and $\boldsymbol{B}$

maximized – whereas for an angle of 90 degrees, i.e. when $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, the dot product is zero.

It is worth noting that the dot product obeys the *commutative law*, that is, $\boldsymbol{A} \bullet \boldsymbol{B} = \boldsymbol{B} \bullet \boldsymbol{A}$.

The *cross product* of two vectors is defined as

$$\begin{aligned} \boldsymbol{A} \times \boldsymbol{B} &= \hat{\boldsymbol{n}} \, |\boldsymbol{A}| \, |\boldsymbol{B}| \sin\theta = \boldsymbol{C} \\ &= \hat{\boldsymbol{x}}(A_y B_z - A_z B_y) + \hat{\boldsymbol{y}}(A_z B_x - A_x B_z) + \hat{\boldsymbol{z}}(A_x B_y - A_y B_x) \end{aligned} \tag{1.9}$$

where $\hat{\boldsymbol{n}}$ is a unit vector normal to the plane containing $\boldsymbol{A}$ and $\boldsymbol{B}$. The cross $\times$ between $\boldsymbol{A}$ and $\boldsymbol{B}$ indicates the cross product, which results in a vector $\boldsymbol{C}$; thus, it is also called a *vector product*. The vector $\boldsymbol{C}$ is orthogonal to both $\boldsymbol{A}$ and $\boldsymbol{B}$, and the direction of $\boldsymbol{C}$ follows a so-called right-hand rule, as shown in Figure 1.9. If the angle θ is zero or 180 degrees, that is, $\boldsymbol{A}$ and $\boldsymbol{B}$ are in parallel, the cross product is zero; whereas for an angle of 90 degrees, i.e. $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, the cross product of these two vectors reaches a maximum. Unlike the dot product, the cross product does not obey the commutative law.

The cross product may be expressed in determinant form as follows, which is the same as Equation (1.9) but may be easier for some people to memorize:

$$\boldsymbol{A} \times \boldsymbol{B} = \begin{vmatrix} \hat{\boldsymbol{x}} & \hat{\boldsymbol{y}} & \hat{\boldsymbol{z}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \tag{1.10}$$

Another important thing about vectors is that any vector can be decomposed into three orthogonal components (such as x, y and z components) in 3D or two orthogonal components in a 2D plane.

Example 1.1: Vector operation. Given vectors $\boldsymbol{A} = 10\hat{\boldsymbol{x}} + 5\hat{\boldsymbol{y}} + 1\hat{\boldsymbol{z}}$ and $\boldsymbol{B} = 2\hat{\boldsymbol{y}}$, find:

$$\boldsymbol{A} + \boldsymbol{B};\ \boldsymbol{A} - \boldsymbol{B};\ \boldsymbol{A} \bullet \boldsymbol{B};\ \text{and}\ \boldsymbol{A} \times \boldsymbol{B}$$

Solution:
$\boldsymbol{A} + \boldsymbol{B} = 10\hat{\boldsymbol{x}} + (5+2)\hat{\boldsymbol{y}} + 1\hat{\boldsymbol{z}} = 10\hat{\boldsymbol{x}} + 7\hat{\boldsymbol{y}} + 1\hat{\boldsymbol{z}}$;
$\boldsymbol{A} - \boldsymbol{B} = 10\hat{\boldsymbol{x}} + (5-2)\hat{\boldsymbol{y}} + 1\hat{\boldsymbol{z}} = 10\hat{\boldsymbol{x}} + 3\hat{\boldsymbol{y}} + 1\hat{\boldsymbol{z}}$;
$\boldsymbol{A} \bullet \boldsymbol{B} = 0 + (5 \times 2) + 0 = 10$;
$\boldsymbol{A} \times \boldsymbol{B} = 10 \times 2\hat{\boldsymbol{z}} + 1 \times 2\hat{\boldsymbol{x}} = 20\hat{\boldsymbol{z}} + 2\hat{\boldsymbol{x}}$

1.3.3 Coordinates

In addition to the well-known Cartesian coordinates, spherical coordinates (r, θ, ϕ), as shown in Figure 1.10, will also be used frequently throughout this book. These two coordinate systems have the following relations:

$$\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned} \tag{1.11}$$

and

$$\begin{aligned} r &= \sqrt{x^2+y^2+z^2} \\ \theta &= \cos^{-1}\frac{z}{\sqrt{x^2+y^2+z^2}};\ 0 \le \theta \le \pi \\ \phi &= \tan^{-1}\frac{y}{x};\ 0 \le \phi \le 2\pi \end{aligned} \tag{1.12}$$

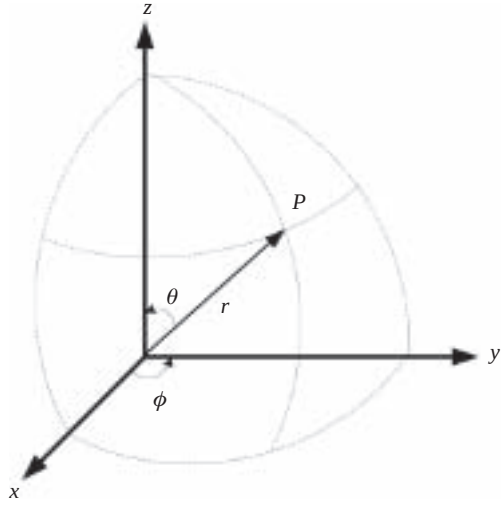


Figure 1.10 Cartesian and spherical coordinates

The dot products of unit vectors in these two coordinate systems are:

$$\begin{aligned}
&\hat{\boldsymbol{x}}\bullet\hat{\boldsymbol{r}} = \sin\theta\cos\phi;\ \hat{\boldsymbol{y}}\bullet\hat{\boldsymbol{r}} = \sin\theta\sin\phi;\ \hat{\boldsymbol{z}}\bullet\hat{\boldsymbol{r}} = \cos\theta \\
&\hat{\boldsymbol{x}}\bullet\hat{\boldsymbol{\theta}} = \cos\theta\cos\phi;\ \hat{\boldsymbol{y}}\bullet\hat{\boldsymbol{\theta}} = \cos\theta\sin\phi;\ \hat{\boldsymbol{z}}\bullet\hat{\boldsymbol{\theta}} = -\sin\theta \\
&\hat{\boldsymbol{x}}\bullet\hat{\phi} = -\sin\phi;\ \hat{\boldsymbol{y}}\bullet\hat{\phi} = \cos\phi;\ \hat{\boldsymbol{z}}\bullet\hat{\phi} = 0
\end{aligned} \tag{1.13}$$

Thus, we can express a quantity in one coordinate system using the known parameters in the other coordinate system. For example, if A_r, A_θ, A_ϕ are known, we can find

$$A_x = \boldsymbol{A}\bullet\hat{\boldsymbol{x}} = A_r\sin\theta\cos\phi + A_\theta\cos\theta\cos\phi - A_\phi\sin\phi$$

1.4 Basics of Electromagnetics

Now let us use basic mathematics to deal with antennas or, more precisely, electromagnetic (EM) problems in this section.

EM waves cover the whole spectrum; radio waves and optical waves are just two examples of EM waves. We can see light but we cannot see radio waves. The whole spectrum is divided into many frequency bands. Some radio frequency bands are listed in Table 1.1.

Although the whole spectrum is infinite, the useful spectrum is limited and some frequency bands, such as the UHF, are already very congested. Normally, significant license fees have to be paid to use the spectrum, although there are some license-free bands: the most well-known ones are the industrial, science and medical (ISM) bands. The 433 MHz and 2.45 GHz are just two examples. Cable operators do not need to pay the spectrum license fees, but they have to pay other fees for things such as digging out the roads to bury the cables.

The *wave velocity*, v, is linked to the frequency, f, and wavelength, λ, by this simple equation:

$$v = \frac{\lambda}{f} \tag{1.14}$$

It is well known that the speed of light (an EM wave) is about 3×10^8 m/s in free space. The higher the frequency, the shorter the wavelength. An illustration of how the frequency is linked

Table 1.1 EM spectrum and applications

Frequency	Band	Wavelength	Applications
3–30 kHz	VLF	100–10 km	Navigation, sonar, fax
30–300 kHz	LF	10–1 km	Navigation
0.3–3 MHz	MF	1–0.1 km	AM broadcasting
3–30 MHz	HF	100–10 m	Tel, fax, CB, ship communications
30–300 MHz	VHF	10–1 m	TV, FM broadcasting
0.3–3 GHz	UHF	1–0.1 m	TV, mobile, radar
3–30 GHz	SHF	100–10 mm	Radar, satellite, mobile, microwave links
30–300 GHz	EHF	10–1 mm	Radar, wireless communications
0.3–3 THz	THz	1–0.1 mm	THz imaging

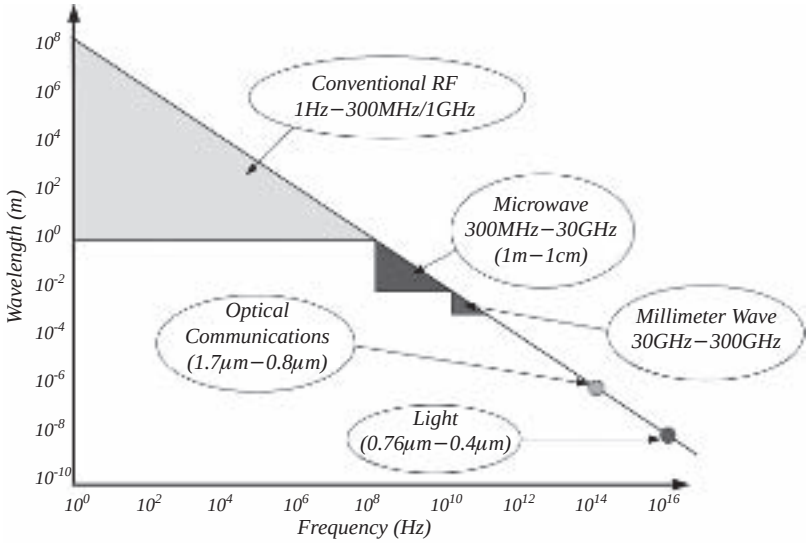


Figure 1.11 Frequency vs wavelength

to the wavelength is given in Figure 1.11, where both the frequency and wavelength are plotted on a logarithmic scale. The advantage of doing this is that we can see clearly how the function is changed, even over a very large scale.

Logarithmic scales are widely used in RF (radio frequency) engineering and the antennas community since the signals we are dealing with change significantly (over 1000 times in many cases) in terms of the magnitude. The signal power is normally expressed in dB and is defined as

$$P(dBW) = 10\log_{10}\frac{P(W)}{1W};\ P(dBm) = 10\log_{10}\frac{P(W)}{1mW} \tag{1.15}$$

Thus, 100 watts is 20 dBW, just expressed as 20 dB in most cases. 1 W is 0 dB or 30 dBm and 0.5 W is −3 dB or 27 dBm. Based on this definition, we can also express other parameters in dB. For example, since the power is linked to voltage V by $P = V^2/R$ (so $P \propto V^2$), the voltage can be converted to dBV by

$$V(dBV) = 20\log_{10}\left(\frac{V(V)}{1V}\right) \tag{1.16}$$

Thus, 3 kVolts is 70 dBV and 0.5 Volts is –6 dBV (not −3 dBV) or 54 dBmV.

1.4.1 The Electric Field

The *electric field* (in V/m) is defined as the force (in Newtons) per unit charge (in Coulombs). From this definition and Coulomb's law, the electric field, $\boldsymbol{E}$, created by a single point

charge Q at a distance r is

$$\boldsymbol{E} = \frac{\boldsymbol{F}}{Q} = \frac{Q}{4\pi\varepsilon r^2}\hat{\boldsymbol{r}}\ (V/m) \tag{1.17}$$

where

$\boldsymbol{F}$ is the electric force given by Coulomb's law ($\boldsymbol{F} = \frac{Q_1 Q_2}{4\pi\varepsilon r^2}\hat{\boldsymbol{r}}$);

$\hat{\boldsymbol{r}}$ is a unit vector along the $\boldsymbol{r}$ direction, which is also the direction of the electric field $\boldsymbol{E}$;

ε is the electric *permittivity* (it is also called the *dielectric constant*, but is normally a function of frequency and not really a constant, thus permittivity is preferred in this book) of the material. Its SI unit is Farads/m. In free space, it is a constant:

$$\varepsilon_0 = 8.85419 \times 10^{-12}\ \mathrm{F/m} \tag{1.18}$$

The product of the permittivity and the electric field is called the *electric flux density*, $\boldsymbol{D}$, which is a measure of how much electric flux passes through a unit area, i.e.

$$\boldsymbol{D} = \varepsilon\boldsymbol{E} = \varepsilon_r\varepsilon_0\boldsymbol{E}(C/m^2) \tag{1.19}$$

where $\varepsilon_r = \varepsilon/\varepsilon_0$ is called the *relative permittivity* or *relative dielectric constant*. The relative permittivities of some common materials are listed in Table 1.2. Note that they are functions of frequency and temperature. Normally, the higher the frequency, the smaller the permittivity in the radio frequency band. It should also be pointed out that almost all conductors have a relative permittivity of one.

The electric flux density is also called the *electric displacement*, hence the symbol $\boldsymbol{D}$. It is also a vector. In an isotropic material (properties independent of direction), $\boldsymbol{D}$ and $\boldsymbol{E}$ are in the same direction and ε is a scalar quantity. In an anisotropic material, $\boldsymbol{D}$ and $\boldsymbol{E}$ may be in different directions if ε is a tensor.

If the permittivity is a complex number, it means that the material has some loss. The *complex permittivity* can be written as

$$\varepsilon = \varepsilon' - j\varepsilon'' \tag{1.20}$$

The ratio of the imaginary part to the real part is called the *loss tangent*, that is

$$\tan\delta = \frac{\varepsilon''}{\varepsilon'} \tag{1.21}$$

It has no unit and is also a function of frequency and temperature.

The electric field $\boldsymbol{E}$ is related to the current density $\boldsymbol{J}$ (in A/m^2), another important parameter, by Ohm's law. The relationship between them at a point can be expressed as

$$\boldsymbol{J} = \sigma\boldsymbol{E} \tag{1.22}$$

where σ is the *conductivity*, which is the reciprocal of *resistivity*. It is a measure of a material's ability to conduct an electrical current and is expressed in Siemens per meter (S/m). Table 1.3

Table 1.2 Relative permittivity of some common materials at 100 MHz

Material	Relative permittivity	Material	Relative permittivity
ABS (plastic)	2.4–3.8	Polypropylene	2.2
Air	1	Polyvinylchloride (PVC)	3
Alumina	9.8	Porcelain	5.1–5.9
Aluminum silicate	5.3–5.5	PTFE-teflon	2.1
Balsa wood	1.37 @ 1 MHz 1.22 @ 3 GHz	PTFE-ceramic	10.2
Concrete	∼8	PTFE-glass	2.1–2.55
Copper	1	RT/Duroid 5870	2.33
Diamond	5.5–10	RT/Duroid 6006	6.15 @ 3 GHz
Epoxy (FR4)	4.4	Rubber	3.0–4.0
Epoxy glass PCB	5.2	Sapphire	9.4
Ethyl alcohol (absolute)	24.5 @ 1 MHz 6.5 @ 3 GHz	Sea water	80
FR-4(G-10)			
–low resin	4.9	Silicon	11.7–12.9
–high resin	4.2		
GaAs	13.0	Soil	∼10
Glass	∼4	Soil (dry sandy)	2.59 @ 1 MHz 2.55 @ 3 GHz
Gold	1	Water (32°F)	88.0
		(68°F)	80.4
		(212°F)	55.3
Ice (pure distilled water)	4.15 @ 1 MHz 3.2 @ 3 GHz	Wood	∼2

Table 1.3 Conductivities of some common materials at room temperature

Material	Conductivity (S/m)	Material	Conductivity (S/m)
Silver	6.3×10^7	Graphite	$\approx 10^5$
Copper	5.8×10^7	Carbon	$\approx 10^4$
Gold	4.1×10^7	Silicon	$\approx 10^3$
Aluminum	3.5×10^7	Ferrite	$\approx 10^2$
Tungsten	1.8×10^7	Sea water	≈ 5
Zinc	1.7×10^7	Germanium	≈ 2
Brass	1×10^7	Wet soil	≈ 1
Phosphor bronze	1×10^7	Animal blood	0.7
Tin	9×10^6	Animal body	0.3
Lead	5×10^6	Fresh water	$\approx 10^{-2}$
Silicon steel	2×10^6	Dry soil	$\approx 10^{-3}$
Stainless steel	1×10^6	Distilled water	$\approx 10^{-4}$
Mercury	1×10^6	Glass	$\approx 10^{-12}$
Cast iron	$\approx 10^6$	Air	0

lists conductivities of some common materials linked to antenna engineering. The conductivity is also a function of temperature and frequency.

1.4.2 The Magnetic Field

Whilst charges can generate an electric field, currents can generate a magnetic field. The *magnetic field*, $\boldsymbol{H}$ (in A/m), is the vector field which forms closed loops around electric currents or magnets. The magnetic field from a current vector $\boldsymbol{I}$ is given by the Biot–Savart law as

$$\boldsymbol{H} = \frac{\boldsymbol{I} \times \hat{\boldsymbol{r}}}{4\pi r^2} (A/m) \tag{1.23}$$

where

$\hat{\boldsymbol{r}}$ is the unit displacement vector from the current element to the field point and
r is the distance from the current element to the field point.
$\boldsymbol{I}$, $\hat{\boldsymbol{r}}$ and $\boldsymbol{H}$ follow the right-hand rule; that is, $\boldsymbol{H}$ is orthogonal to both $\boldsymbol{I}$ and $\hat{\boldsymbol{r}}$, as illustrated by Figure 1.12.

Like the electric field, the magnetic field exerts a force on electric charge. But unlike an electric field, it employs force only on a moving charge, and the direction of the force is orthogonal to both the magnetic field and the charge's velocity:

$$\boldsymbol{F} = Q\boldsymbol{v} \times \mu \boldsymbol{H} \tag{1.24}$$

where

$\boldsymbol{F}$ is the force vector produced, measured in Newtons;
Q is the electric charge that the magnetic field is acting on, measured in Coulombs (C);
$\boldsymbol{v}$ is the velocity vector of the electric charge Q, measured in meters per second (m/s);
μ is the magnetic permeability of the material. Its unit is Henries per meter (H/m). In free space, the permeability is

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m} \tag{1.25}$$

In Equation (1.24), $Q\boldsymbol{v}$ can actually be viewed as the current vector $\boldsymbol{I}$ and the product of $\mu\boldsymbol{H}$ is called the *magnetic flux density* $\boldsymbol{B}$ (in Tesla), the counterpart of the electric flux density.

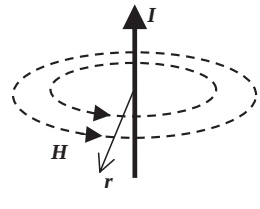


Figure 1.12 Magnetic field generated by current I

Table 1.4 Relative permeabilities of some common materials

Material	Relative permeability	Material	Relative permeability
Superalloy	$\approx 1 \times 10^6$	Aluminum	≈ 1
Purified iron	$\approx 2 \times 10^5$	Air	1
Silicon iron	$\approx 7 \times 10^3$	Water	≈ 1
Iron	$\approx 5 \times 10^3$	Copper	≈ 1
Mild steel	$\approx 2 \times 10^3$	Lead	≈ 1
Nickel	600	Silver	≈ 1

Thus

$$\boldsymbol{B} = \mu \boldsymbol{H} \tag{1.26}$$

Again, in an isotropic material (properties independent of direction), $\boldsymbol{B}$ and $\boldsymbol{H}$ are in the same direction and μ is a scalar quantity. In an anisotropic material, $\boldsymbol{B}$ and $\boldsymbol{E}$ may be in different directions and μ is a tensor.

Like the relative permittivity, the relative permeability is given as

$$\mu_r = \mu/\mu_0 \tag{1.27}$$

The relative permeabilities of some materials are given in Table 1.4. Permeability is not sensitive to frequency or temperature. Most materials, including conductors, have a relative permeability very close to one.

Combining Equations (1.17) and (1.24) yields

$$\boldsymbol{F} = Q(\boldsymbol{E} + \boldsymbol{v} \times \mu \boldsymbol{H}) \tag{1.28}$$

This is called the *Lorentz force*. The particle will experience a force due to the electric field $Q\boldsymbol{E}$, and the magnetic field $Q\boldsymbol{v} \times \boldsymbol{B}$.

1.4.3 Maxwell's Equations

Maxwell's equations are a set of equations first presented as a distinct group in the latter half of the nineteenth century by James Clerk Maxwell (1831–1879), pictured in Figure 1.13. Mathematically they can be expressed in the following differential form:

$$\begin{aligned} \nabla \times \boldsymbol{E} &= -\frac{d\boldsymbol{B}}{dt} \\ \nabla \times \boldsymbol{H} &= \boldsymbol{J} + \frac{d\boldsymbol{D}}{dt} \\ \nabla \bullet \boldsymbol{D} &= \rho \\ \nabla \bullet \boldsymbol{B} &= 0 \end{aligned} \tag{1.29}$$

where

ρ is the charge density;

$\nabla = \frac{\partial}{\partial x}\hat{\boldsymbol{x}} + \frac{\partial}{\partial y}\hat{\boldsymbol{y}} + \frac{\partial}{\partial z}\hat{\boldsymbol{z}}$ is a vector operator;

Figure 1.13 James Clerk Maxwell

$\nabla\times$ is the *curl operator*, called *rot* in some countries instead of *curl*;
$\nabla\bullet$ is the *divergence operator*.

Here we have both the vector cross product and dot product.

Maxwell's equations describe the interrelationship between electric fields, magnetic fields, electric charge and electric current. Although Maxwell himself was not the originator of the individual equations, he derived them again independently in conjunction with his molecular vortex model of Faraday's lines of force, and he was the person who first grouped these equations together into a coherent set. Most importantly, he introduced an extra term to Ampere's Circuital Law, the second equation of (1.19). This extra term is the time derivative of the electric field and is known as *Maxwell's displacement current*. Maxwell's modified version of Ampere's Circuital Law enables the set of equations to be combined together to derive the electromagnetic wave equation, which will be further discussed in Chapter 3.

Now let us have a closer look at these mathematical equations to see what they really mean in terms of the physical explanations.

1.4.3.1 Faraday's Law of Induction

$$\nabla \times \boldsymbol{E} = \frac{d\boldsymbol{B}}{dt} \tag{1.30}$$

This equation simply means that the induced electromotive force is proportional to the rate of change of the magnetic flux through a coil. In layman's terms, moving a conductor (such as a metal wire) through a magnetic field produces a voltage. The resulting voltage is directly proportional to the speed of movement. It is apparent from this equation that a time-varying magnetic field ($\mu\frac{d\boldsymbol{H}}{dt} \neq 0$) will generate an electric field, i.e. $\boldsymbol{E} \neq 0$. But if the magnetic field is not time-varying, it will NOT generate an electric field.

1.4.3.2 Ampere's Circuital Law

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{d\boldsymbol{D}}{dt} \tag{1.31}$$

This equation was modified by Maxwell by introducing the displacement current $\frac{d\boldsymbol{D}}{dt}$. It means that a magnetic field appears during the charge or discharge of a capacitor. With this concept, and Faraday's law, Maxwell was able to derive the wave equations, and by showing that the predicted wave velocity was the same as the measured velocity of light, Maxwell asserted that light waves are electromagnetic waves.

This equation shows that both the current ($\boldsymbol{J}$) and time-varying electric field ($\varepsilon \frac{d\boldsymbol{E}}{dt}$) can generate a magnetic field, i.e. $\boldsymbol{H} \neq 0$.

1.4.3.3 Gauss's Law for Electric Fields

$$\nabla \bullet \boldsymbol{D} = \rho \tag{1.32}$$

This is the electrostatic application of Gauss's generalized theorem, giving the equivalence relation between any flux, e.g. of liquids, electric or gravitational, flowing out of any closed surface and the result of inner sources and sinks, such as electric charges or masses enclosed within the closed surface. As a result, it is not possible for electric fields to form a closed loop. Since $\boldsymbol{D} = \varepsilon \boldsymbol{E}$, it is also clear that charges (ρ) can generate electric fields, i.e. $\boldsymbol{E} \neq 0$.

1.4.3.4 Gauss's Law for Magnetic Fields

$$\nabla \bullet \boldsymbol{B} = 0 \tag{1.33}$$

This shows that the divergence of the magnetic field ($\nabla \bullet \mathbf{B}$) is always zero, which means that the magnetic field lines are closed loops; thus, the integral of $\boldsymbol{B}$ over a closed surface is zero.

For a time-harmonic electromagnetic field (which means a field linked to time by factor $e^{j\omega t}$ where ω is the angular frequency and t is the time), we can use the *constitutive relations*

$$\boldsymbol{D} = \varepsilon \boldsymbol{E}, \quad \boldsymbol{B} = \mu \boldsymbol{H}, \quad \boldsymbol{J} = \sigma \boldsymbol{E} \tag{1.34}$$

to write Maxwell's equations in the following form

$$\begin{aligned} &\nabla \times \boldsymbol{E} = -j\omega\mu \boldsymbol{H} \\ &\nabla \times \boldsymbol{H} = \boldsymbol{J} + j\omega\varepsilon \boldsymbol{E} = j\omega\varepsilon \left(1 - j\frac{\sigma}{\omega\varepsilon}\right) \boldsymbol{E} \\ &\nabla \bullet \boldsymbol{E} = \rho/\varepsilon \\ &\nabla \bullet \boldsymbol{H} = 0 \end{aligned} \tag{1.35}$$

where $\boldsymbol{B}$ and $\boldsymbol{D}$ are replaced by the electric field $\boldsymbol{E}$ and magnetic field $\boldsymbol{H}$ to simplify the equations and they will not appear again unless necessary.

It should be pointed out that, in Equation (1.35), $\varepsilon(1 - j\frac{\sigma}{\omega\varepsilon})$ can be viewed as a *complex permittivity* defined by Equation (1.20). In this case, the loss tangent is

$$\tan\delta = \frac{\varepsilon''}{\varepsilon'} = \frac{\sigma}{\omega\varepsilon} \tag{1.36}$$

It is hard to predict how the loss tangent changes with the frequency, since both the permittivity and conductivity are functions of frequency as well. More discussion will be given in Chapter 3.

1.4.4 Boundary Conditions

Maxwell's equations can also be written in the integral form as

$$\begin{aligned}
&\oint_C \boldsymbol{E}\bullet d\boldsymbol{l} = -\iint_S \frac{d\boldsymbol{B}}{dt}\bullet d\boldsymbol{s} \\
&\oint_C \boldsymbol{H}\bullet d\boldsymbol{l} = \iint_S (\boldsymbol{J} + \frac{d\boldsymbol{D}}{dt})\bullet d\boldsymbol{s} \\
&\oiint_S \boldsymbol{D}\bullet d\boldsymbol{s} = \iiint_V \rho dv = Q \\
&\oiint_S \boldsymbol{B}\bullet d\boldsymbol{s} = 0
\end{aligned} \tag{1.37}$$

Consider the boundary between two materials shown in Figure 1.14. Using these equations, we can obtain a number of useful results. For example, if we apply the first equation of Maxwell's equations in integral form to the boundary between Medium 1 and Medium 2, it is not difficult to obtain [2]:

$$\hat{\boldsymbol{n}} \times \boldsymbol{E_1} = \hat{\boldsymbol{n}} \times \boldsymbol{E}_2 \tag{1.38}$$

where $\hat{\boldsymbol{n}}$ is the surface unit vector from Medium 2 to Medium 1, as shown in Figure 1.14. This condition means that the tangential components of an electric field ($\hat{\boldsymbol{n}} \times \boldsymbol{E}$) are continuous across the boundary between any two media.

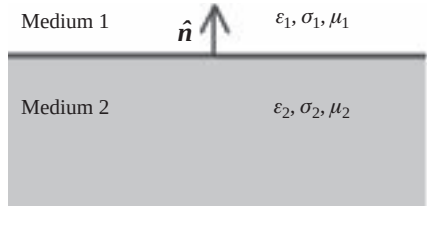


Figure 1.14 Boundary between Medium 1 and Medium 2

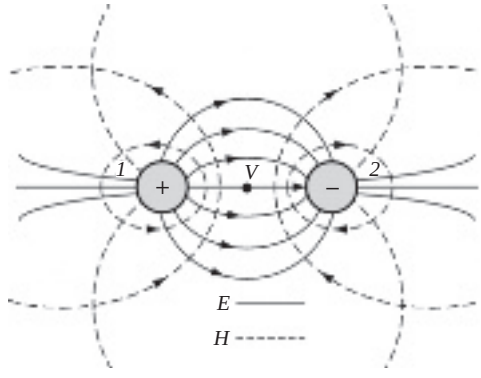


Figure 1.15 Electromagnetic field distribution around a two-wire transmission line

Similarly, we can apply the other three Maxwell equations to this boundary to obtain:

$$\begin{aligned} &\hat{\boldsymbol{n}} \times (\boldsymbol{H_1} - \boldsymbol{H_2}) = \boldsymbol{J_s} \\ &\hat{\boldsymbol{n}} \bullet (\varepsilon_1 \boldsymbol{E_1} - \varepsilon_2 \boldsymbol{E_2}) = \rho_s \\ &\hat{\boldsymbol{n}} \bullet (\mu_1 \boldsymbol{H_1} - \mu_2 \boldsymbol{H_2}) = 0 \end{aligned} \tag{1.39}$$

where $\boldsymbol{J}_s$ is the surface current density and ρ_s is the surface charge density. These results can be interpreted as

- the change in tangential component of the magnetic field across a boundary is equal to the surface current density on the boundary;
- the change in the normal component of the electric flux density across a boundary is equal to the surface charge density on the boundary;
- the normal component of the magnetic flux density is continuous across the boundary between two media, whilst the normal component of the magnetic field is not continuous unless $\mu_1 = \mu_2$.

Applying these boundary conditions on a perfect conductor (which means no electric and magnetic field inside and the conductivity $\sigma = \infty$) in the air, we have

$$\hat{\boldsymbol{n}} \times \boldsymbol{E} = 0;\ \hat{\boldsymbol{n}} \times \boldsymbol{H} = \boldsymbol{J_s};\ \hat{\boldsymbol{n}} \bullet \boldsymbol{E} = \rho_s / \varepsilon;\ \hat{\boldsymbol{n}} \bullet \boldsymbol{H} = 0 \tag{1.40}$$

We can also use these results to illustrate, for example, the field distribution around a two-wire transmission line, as shown in Figure 1.15, where the electric fields are plotted as the solid lines and the magnetic fields are shown as broken lines. As expected, the electric field is from positive charges to negative charges, whilst the magnetic field forms loops around the current.

1.5 Summary

In this chapter we have introduced the concept of antennas, briefly reviewed antenna history and laid down the mathematical foundations for further study. The focus has been on the basics of electromagnetics, which include electric and magnetic fields, electromagnetic properties of materials, Maxwell's equations and boundary conditions. Maxwell's equations have revealed how electric fields, magnetic fields and sources (currents and charges) are interlinked. They are the foundation of electromagnetics and antennas.

References

[1] R. E. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, Inc., 1985.
[2] J. D. Kraus and D. A. Fleisch, *Electromagnetics with Applications*, 5th edition, McGraw-Hill, Inc., 1999.

Problems

Q1.1 What wireless communication experiment did H. Hertz conduct in 1887? Use a diagram to illustrate your answer.
Q1.2 Use an example to explain what a complex number means in our daily life.
Q1.3 Given vectors $\boldsymbol{A} = 10\hat{\boldsymbol{x}} + 5\hat{\boldsymbol{y}} + 1\hat{\boldsymbol{z}}$ and $\boldsymbol{B} = 5\hat{\boldsymbol{z}}$, find
a. the amplitude of vector $\boldsymbol{A}$;
b. the angle between vectors $\boldsymbol{A}$ and $\boldsymbol{B}$;
c. the dot product of these two vectors;
d. a vector which is orthogonal to $\boldsymbol{A}$ and $\boldsymbol{B}$.
Q1.4 Given vector $\boldsymbol{A} = 10\sin(10t + 10z)\hat{\boldsymbol{x}} + 5\hat{\boldsymbol{y}}$, find
a. $\nabla \bullet \boldsymbol{A}$;
b. $\nabla \times \boldsymbol{A}$;
c. $(\nabla \bullet \nabla)\,\boldsymbol{A}$;
d. $\nabla\nabla \bullet \boldsymbol{A}$
Q1.5 Vector $\boldsymbol{E} = 10e^{j(10t-10z)}\hat{\boldsymbol{x}}$.
a. find the amplitude of $\boldsymbol{E}$;
b. plot the real part of $\boldsymbol{E}$ as a function of t;
c. plot the real part of $\boldsymbol{E}$ as a function of z;
d. explain what this vector means.
Q1.6 Explain why mobile phone service providers have to pay license fees to use the spectrum. Who is responsible for the spectrum allocation in your country?
Q1.7 Cellular mobile communications have become part of our daily life. Explain the major differences between the 1st, 2nd and 3rd generations of cellular mobile systems in terms of the frequency, data rate and bandwidth. Further explain why their operational frequencies have increased.
Q1.8 Which frequency bands have been used for radar applications? Give an example.
Q1.9 Express 1 kW in dB, 10 kV in dBV, 0.5 dB in W and 40 dBμV/m in V/m and μV/m.
Q1.10 Explain the concepts of the electric field and magnetic field. How are they linked to the electric and magnetic flux density functions?

Q1.11 What are the material properties of interest to our electromagnetic and antenna engineers?

Q1.12 What is the Lorentz force? Name an application of the Lorentz force in our daily life.

Q1.13 If a magnetic field on a perfect conducting surface $z = 0$ is $\boldsymbol{H} = 10\cos(10t - 5z)\hat{\boldsymbol{x}}$, find the surface current density $\boldsymbol{J}_s$.

Q1.14 Use Maxwell's equations to explain the major differences between static EM fields and time-varying EM fields.

Q1.15 Express the boundary conditions for the electric and magnetic fields on the surface of a perfect conductor.

2

Circuit Concepts and Transmission Lines

In this chapter we are going to review the very basics of circuit concepts and distinguish the lumped element system from the distributed element system. The focus will be on the fundamentals of transmission lines, including the basic model, the characteristic impedance, input impedance, reflection coefficient, return loss and voltage standing wave ratio (VSWR) of a transmission line. The Smith Chart, impedance-matching techniques, Q factor and bandwidth will also be addressed. A comparison of various transmission lines and associated connectors will be made at the end of this chapter.

2.1 Circuit Concepts

Figure 2.1 shows a very basic electrical circuit where a voltage source V is connected to a load Z via conducting wires. This simple circuit can represent numerous systems in our daily life, from a simple torch – a DC (direct current) circuit – to a more complicated power supply system – an AC (alternating current) circuit. To analyze such a circuit, one has to use the following four quantities:

- *Electric current* I is a measure of the charge flow/movement. The SI unit of current is the Ampere (A), which is equal to a flow of one Coulomb of charge per second.
- *Voltage* V is the difference in electrical potential between two points of an electrical or electronic circuit. The SI unit of voltage is the Volt (V). Voltage measures the potential energy of an electric field to cause an electric current in a circuit.
- *Impedance* $Z = R + jX$ is a measure of opposition to an electric current. In general, the impedance is a complex number, its real part R is the electrical *resistance* (or just resistance) and reflects the ability to consume energy, whilst the imaginary part X is the *reactance* and indicates the ability to store energy. If the reactance is positive, it is called *inductance* since the reactance of an inductor is positive (ωL); if the reactance is negative, it is then

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

Figure 2.1 A simple electrical circuit with a source and load

called *capacitance* since the reactance of a capacitor is negative ($-1/\omega C$). The same unit, the Ohm (Ω), is used for impedance, resistance and reactance. The inverses of the impedance, resistance and reactance are called the *admittance* (Y), *conductance* (G) and *susceptance* (B), respectively. Their unit is the Siemens (S) and it is 1 Ohm.

- *Power P* is defined as the amount of work done by an electrical current, or the rate at which electrical energy is transmitted/consumed. The SI unit of power is the Watt (W). When an electric current flows through a device with resistance, the device converts the power into various forms, such as light (light bulbs), heat (electric cooker), motion (electric razor), sound (loudspeaker) or radiation for an antenna.

Ohm's law is the fundamental theory for electrical circuits. It reveals how the current, voltage and resistance are linked in a DC circuit. It states that the current passing through a conductor/device from one terminal point on the conductor/device to another terminal point on the conductor/device is directly proportional to the potential difference (i.e. voltage) across the two terminal points and inversely proportional to the resistance of the conductor/device between the two terminal points. That is

$$I = \frac{V}{R} \tag{2.1}$$

In an AC circuit, Ohm's law can be generalized as

$$I = \frac{V}{Z} \tag{2.2}$$

i.e. the resistance R is replaced by the impedance Z. Since the impedance is a complex number, both the current and voltage can be complex numbers as well, which means that they have magnitude and phase.

The average power can be obtained using

$$P = IV = V^2/R = RI^2 \quad \text{for DC} \tag{2.3}$$

$$P_{av} = \frac{1}{2} I_0 V_0 = \frac{V_0^2}{2R} = \frac{1}{2} R I_0^2 \quad \text{for AC}$$

where V_0 and I_0 are the amplitudes of voltage and current, respectively.

2.1.1 Lumped and Distributed Element Systems

In traditional circuit theory, we basically divide circuits into those that are DC and those that are AC. The voltage, current and impedance are real numbers in DC circuits but complex numbers in AC circuits. The effects of conducting wires can normally be neglected. For example, the current across the load Z in Figure 2.1 can be obtained using Ohm's law. It is given by Equation (2.2) and considered to be the same voltage across the load.

In most countries, the electrical power supply system operates at 50 or 60 Hz, which means a wavelength of 6000 or 5000 km (close to the radius of the Earth: 6378 km), much longer than any transmission line in use. The current and voltage along the transmission line may be considered unchanged. The system is called a *lumped element system*. However, in some applications the frequency of the source is significantly increased, as a result the wavelength becomes comparable with the length of the transmission line linking the source and the load. The current and voltage along the transmission line are functions of the distance from the source, thus the system is called a *distributed element system*. If Figure 2.1 is a distributed element system, Equation (2.2) is no longer valid since the voltage across the load may now be very different from the source voltage. V_0 should therefore be replaced by the voltage across the load.

Conventional circuit theory was developed for lumped element systems whose frequency is relatively low and where the wavelength is relatively large. However, the frequency of a distributed system is relatively high and the wavelength is relatively short. It is therefore important to introduce the transmission line theory, which has been developed for the distributed element system and has taken the distributed nature of the parameters in the system into account.

2.2 Transmission Line Theory

A *transmission line* is the structure that forms all or part of a path from one place to another for directing the transmission of energy, such as electrical power transmission and optical waves. Examples of transmission lines include conducting wires, electrical power lines, coaxial cables, dielectric slabs, optical fibers and waveguides. In this book we are only interested in the transmission lines for RF engineering and antenna applications. Thus, dielectric transmission lines such as optical fibers are not considered.

2.2.1 Transmission Line Model

The simplest transmission line is a two-wire conducting transmission line, as shown in Figure 2.2. It has been widely used for electrical power supply and also for radio and television systems. In the old days, the broadcasting TV signal was received by an antenna and then passed down to a TV via such a two-wire conducting wire, which has now been replaced by the coaxial cable. This is partially due to the fact that the antenna used now (the Yagi–Uda antenna, a popular TV antenna to be discussed in Chapter 5, which has an input impedance around 75 ohms) is different from the antenna used then (the folded dipole, which was a popular TV antenna many years ago and had an input impedance around 300 ohms). Also, the coaxial cable performs much better than the two-wire transmission line at the UHF (ultra-high frequency) TV bands.

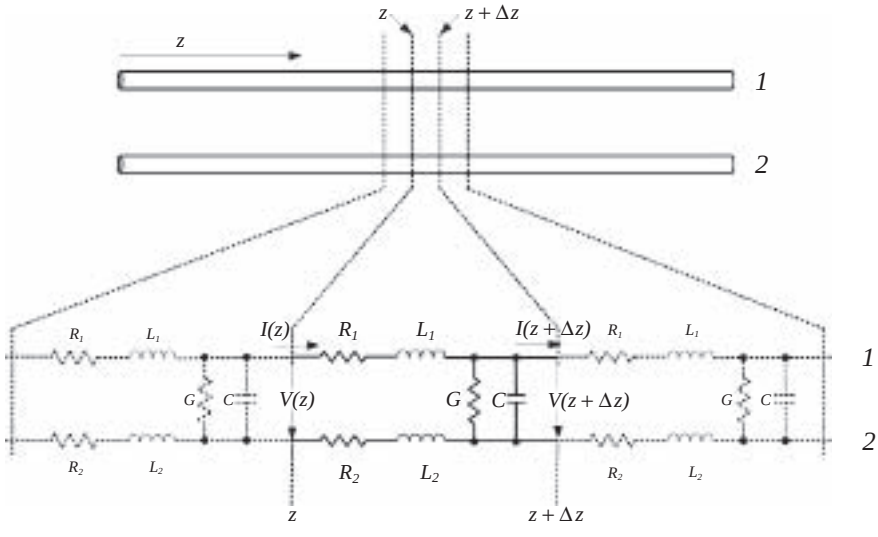


Figure 2.2 A two-wire transmission line model

As shown in Figure 2.2, if we divide the transmission line into many (almost infinite) short segments of length Δz, which is much smaller than the wavelength of interest, each segment can then be represented using a set of lumped elements. By doing so, a distributed transmission line is modeled as an infinite series of two-port lumped elementary components, each representing an infinitesimally short segment of the transmission line. To make the analysis easier, the equivalent circuit of the segment of the transmission line is simplified to Figure 2.3, where $R = R_1 + R_2$ and $L = L_1 + L_2$.

- The resistance R represents the conductive loss of the transmission line over a unit length, thus the unit is ohms/unit length (Ω/m).
- The inductance L is the self-inductance of the transmission line and is expressed in Henries per unit length (H/m).
- The capacitance C between the two conductors is represented by a shunt capacitor with a unit of Farads per unit length (F/m).
- The conductance G of the dielectric material separating the two conductors is represented by a conductance G shunted between the two conductors. Its unit is Siemens per unit length (S/m).

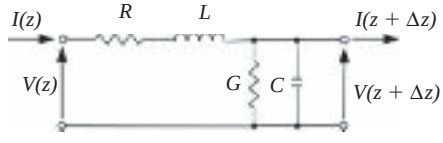


Figure 2.3 Schematic representation of the elementary component of a transmission line

It should be repeated for clarity that the model consists of an infinite series of elements shown in Figure 2.3, and that the values of the components are specified per unit length. R, L, C and G may be functions of frequency.

Using this model we are going to investigate how the current and voltage along the line are changed and how they are linked to R, L, C and G.

It is reasonable to assume that the source is time-harmonic and has an angular frequency $\omega(= 2\pi f$, where f is the frequency), thus its time factor is $e^{j\omega t}$.

Using Ohm's law, we know that the voltage drop and current change over a segment of Δz can be expressed in the frequency domain as:

$$\begin{aligned} V(z+\Delta z) - V(z) &= -(R + j\omega L)\Delta z \cdot I(z) \\ I(z+\Delta z) - I(z) &= -(G + j\omega C)\Delta z \cdot V(z+\Delta z) \end{aligned} \tag{2.4}$$

When Δz approaches zero, these two equations can be written in differential form as:

$$\begin{aligned} \frac{dV(z)}{dz} &= -(R + j\omega L) \cdot I(z) \\ \frac{dI(z)}{dz} &= -(G + j\omega C) \cdot V(z) \end{aligned} \tag{2.5}$$

Differentiating with respect to z on both sides of the equations and combining them gives:

$$\begin{aligned} \frac{d^2V(z)}{dz^2} &= (R + j\omega L)(G + j\omega C) \cdot V(z) \\ \frac{d^2I(z)}{dz^2} &= (R + j\omega L)(G + j\omega C) \cdot I(z) \end{aligned} \tag{2.6}$$

That is,

$$\begin{aligned} \frac{d^2V(z)}{dz^2} - \gamma^2 V(z) &= 0 \\ \frac{d^2I(z)}{dz^2} - \gamma^2 I(z) &= 0 \end{aligned} \tag{2.7}$$

where

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{2.8}$$

and is called the *propagation constant*, which may have real and imaginary parts. Equation (2.7) is a pair of linear differential equations which describe the line voltage and current on a transmission line as a function of distance and time (the time factor $e^{j\omega t}$ is omitted here). They are called *telegraph equations* or *transmission line equations*.

2.2.2 Solutions and Analysis

The general solution of $V(z)$ in the telegraph equations can be expressed as the sum of the forward and reverse voltages [1, 2]

$$V(z) = V_+(z) + V_-(z) = A_1 e^{-\gamma z} + A_2 e^{\gamma z} \tag{2.9}$$

where A_1 and A_2 are complex coefficients to be determined by the boundary conditions, which means the voltage, current and impedance at the input and the load of the transmission line – we need to know at least two of these in order to determine the two coefficients.

Replacing $V(z)$ in Equation (2.5) by Equation (2.9), we can find the solution of the line current as

$$I(z) = \frac{\gamma}{R + j\omega L}(A_1 e^{-\gamma z} - A_2 e^{\gamma z}) \tag{2.10}$$

This can be written as

$$I(z) = \frac{1}{Z_0}(A_1 e^{-\gamma z} - A_2 e^{\gamma z}) \tag{2.11}$$

where

$$Z_0 = \frac{V_+(z)}{I_+(z)} = \frac{R + j\omega L}{\gamma} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{2.12}$$

and is the ratio of the forward voltage to the current thus it is called the *characteristic impedance* of the transmission line. Its unit is the ohm (Ω). It is a function of the frequency and parameters of the line. The industrial standard transmission line normally has a characteristic impedance of 50 or 75 Ω when the loss can be neglected ($R \approx 0$ and $G \approx 0$).

Since the propagation constant is complex, it can be written as:

$$\gamma = \alpha + j\beta \tag{2.13}$$

where α is called the *attenuation constant* (in Nepers/meter, or Np/m) and β is called the *phase constant*. Because $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$, we can find that mathematically:

$$\begin{aligned} \alpha &= \left[\frac{1}{2}\left(\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} + (RG - \omega^2 LC)\right)\right]^{1/2} \\ \beta &= \left[\frac{1}{2}\left(\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} - (RG - \omega^2 LC)\right)\right]^{1/2} \end{aligned} \tag{2.14}$$

They are functions of frequency as well as the parameters of the transmission line.

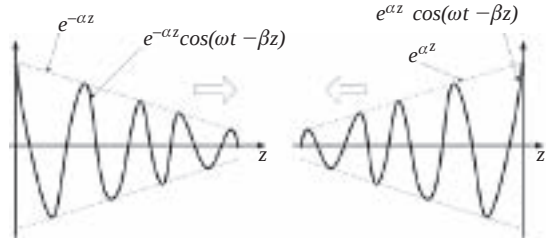


Figure 2.4 Forward- and reverse-traveling waves

If we take the time factor into account, the complete solution of the voltage and current along a transmission line can be expressed as

$$\begin{aligned} V(z,t) &= A_1 e^{j\omega t-\gamma z} + A_2 e^{j\omega t+\gamma z} = A_1 e^{-\alpha z+j(\omega t-\beta z)} + A_2 e^{\alpha z+j(\omega t+\beta z)} \\ I(z,t) &= \tfrac{1}{Z_0}(A_1 e^{j\omega t-\gamma z} - A_2 e^{j\omega t+\gamma z}) = \tfrac{1}{Z_0}(A_1 e^{-\alpha z+j(\omega t-\beta z)} - A_2 e^{\alpha z+j(\omega t+\beta z)}) \end{aligned} \tag{2.15}$$

Physically, the line voltage solution can be considered the combination of two traveling voltage waves: the wave traveling towards the z direction (called the forward wave) has an amplitude of $|V_+(z)| = \left|A_1 e^{-\alpha z}\right|$, which attenuates as z increases, whereas the wave traveling towards the $-z$ direction (called the reverse wave) has an amplitude of $|A_2 e^{\alpha z}|$, as shown in Figure 2.4. The amplitudes of A_1 and A_2 are actually the voltage amplitudes of the forward and reverse waves at $z = 0$, respectively. If there is no reflection at the end of the transmission line, it means that the boundary conditions have forced A_2 to be zero, thus the reverse wave will be zero and only the forward-traveling voltage will exist on the transmission line in this case.

Similarly, the line current can also be viewed as the combination of two traveling current waves. It is worth noting that the reverse-traveling current has a minus sign with the amplitude, this means a phase change of 180 degrees and reflects the direction change in the returned current.

The velocity of the wave is another parameter of interest and it can be determined from the phase term: $\omega t - \beta z$. At a fixed reference point, the wave moves Δz over a period of Δt, i.e. we have $\omega \Delta t - \beta \Delta z = 0$, thus the velocity

$$v = \frac{dz}{dt} = \frac{\omega}{\beta} \tag{2.16}$$

Since the phase constant β is a function of the angular frequency, as shown in Equation (2.14), the velocity is a function of frequency, which is a well-known dispersion problem (change with frequency).

Using Equation (2.16), the phase constant can be expressed as

$$\beta = \frac{\omega}{v} = \frac{2\pi f}{v} = \frac{2\pi}{\lambda} \tag{2.17}$$

where λ is the wavelength. The phase constant is also called the *wave number*. For every one wavelength, the phase is changed by 2π.

These solutions are general and can be applied to any transmission line in principle. We can see that the characteristic impedance may be complex and the attenuation constant and phase constant are complicated functions of frequency. But, in practice, we always prefer something simpler and easier to use.

2.2.2.1 Lossless Transmission Lines

Since the function of the transmission line is to transmit information from one place to another with little change, the loss of the transmission line should be minimized – this is one of the requirements for transmission line manufacture. There are indeed many low-loss transmission lines available on the market.

For a lossless transmission line, elements R and G can be considered to be zero ($R \approx 0$ and $G \approx 0$). In this hypothetical case, the model depends only on elements L and C, which greatly simplifies the analysis.

The characteristic impedance of the transmission line, Equation (2.12), can now be simplified to

$$Z_0 = \sqrt{\frac{L}{C}} \tag{2.18}$$

This is just a real number (resistance) and is determined only by L and C; it is not a function of the frequency.

Similarly, Equation (2.14) becomes

$$\begin{aligned} \alpha &= 0 \\ \beta &= \omega\sqrt{LC} \end{aligned} \tag{2.19}$$

This means that there is no attenuation and the propagation constant is now just an imaginary number:

$$\gamma = j\beta = j\omega\sqrt{LC}$$

The voltage and current along the line are

$$\begin{aligned} V(z,t) &= A_1 e^{j(\omega t - \beta z)} + A_2 e^{j(\omega t + \beta z)} \\ I(z,t) &= \tfrac{1}{Z_0}(A_1 e^{j(\omega t - \beta z)} - A_2 e^{j(\omega t + \beta z)}) \end{aligned} \tag{2.20}$$

Neither the forward nor the reverse wave is attenuated and their amplitudes are not a function of the distance.

The velocity of the waves is now

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{2.21}$$

which is not a function of the frequency and is only determined by the transmission line itself–no dispersion. This is an important feature required for all transmission lines.

2.2.2.2 Low-Loss Transmission Lines

In practice, most transmission lines cannot be considered lossless structures but, instead, are low-loss transmission lines.

The definition of 'low loss' is

$$R \ll \omega L, \qquad G \ll \omega C$$

This seems to imply extremely high frequency. The reality is that both R and G are functions of frequency. Normally, the higher the frequency, the larger R and G. Thus, this condition does not mean high frequency. It applies for any frequency when this condition is met.

For a low-loss transmission line, the characteristic impedance is

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt{\frac{j\omega L(1 + R/j\omega L)}{j\omega C(1 + G/j\omega C)}} \approx \sqrt{\frac{j\omega L(1+0)}{j\omega C(1+0)}} = \sqrt{\frac{L}{C}} \tag{2.22}$$

Thus, it is the same as the lossless case. The characteristic impedance is a pure resistance and is determined by L and C; it is not a function of the frequency. This is why the characteristic impedance of industrial standard transmissions has a constant value, normally 50 or 75 ohms, over a large frequency band, even when the line loss is not zero.

Similarly, the attenuation and phase constants can be approximated as

$$\begin{aligned} \alpha &\approx \frac{R}{2}\sqrt{C/L} + \frac{G}{2}\sqrt{L/C} = \frac{R}{2Z_0} + \frac{GZ_0}{2} \\ \beta &\approx \omega\sqrt{LC} \end{aligned} \tag{2.23}$$

The loss (attenuation) is caused by the resistive loss R and the material loss G between the conductors. The phase constant is again the same as for lossless lines.

However, the voltage and current are attenuated as they travel along the line (due to the loss) and can be expressed as

$$\begin{aligned} V(z,t) &= A_1 e^{-\alpha z + j(\omega t - \beta z)} + A_2 e^{\alpha z + j(\omega t + \beta z)} \\ I(z,t) &= \frac{1}{Z_0}(A_1 e^{-\alpha z + j(\omega t - \beta z)} - A_2 e^{\alpha z + j(\omega t + \beta z)}) \end{aligned} \tag{2.24}$$

where the attenuation constant is given by Equation (2.23).

Just like the lossless line, the velocity of the waves in a low-loss transmission line is only determined by L and C, i.e.

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{2.25}$$

There is no dispersion (it is not changed with frequency).

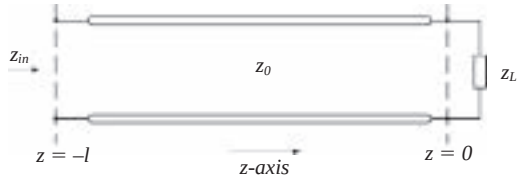


Figure 2.5 A transmission line terminated with a load

2.2.3 Terminated Transmission Line

It is now clear that the voltage and current of a transmission line are distributed quantities; they are functions of the position z. However, the characteristic impedance of a transmission line is not a distributed parameter but a constant. When the line is terminated with a load impedance Z_L, as shown in Figure 2.5, what is the input impedance?

2.2.3.1 Input Impedance

The *input impedance* of a transmission line is defined as the ratio of voltage to current at the input port and is the impedance looking towards the load, i.e.

$$Z_{in}(z) = \frac{V(z)}{I(z)} = Z_0 \frac{A_1 e^{-\gamma z} + A_2 e^{\gamma z}}{A_1 e^{-\gamma z} - A_2 e^{\gamma z}} = Z_0 \frac{e^{-\gamma z} + \Gamma_0 e^{\gamma z}}{e^{-\gamma z} - \Gamma_0 e^{\gamma z}} \tag{2.26}$$

where $\Gamma_0 = A_2 / A_1$ is called the *reflection coefficient* at the load and is equal to the ratio of A_2 to $A_{1,}$ which are the complex coefficients of the forward and reverse voltage waves at $z = 0$. The input impedance at the load should be the load impedance, that is

$$Z_{in}(0) = Z_0 \frac{1 + \Gamma_0}{1 - \Gamma_0} = Z_L \tag{2.27}$$

Thus, the reflection coefficient at the load can be expressed as

$$\Gamma_0 = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.28}$$

A general expression of the reflection coefficient on a transmission line at reference point z is

$$\Gamma(z) = \frac{V_-(z)}{V_+(z)} = \frac{A_2 e^{\gamma z}}{A_1 e^{-\gamma z}} = \Gamma_0 e^{2\gamma z} = \frac{Z_L - Z_0}{Z_L + Z_0} e^{2\gamma z} \tag{2.29}$$

This means that the reflection coefficient is a distributed parameter and is a function of the load impedance as well as the transmission line characteristic impedance.

Replacing Γ_0 in Equation (2.26) by Equation (2.28), we have

$$Z_{in}(z) = Z_0 \frac{(Z_L + Z_0)e^{-\gamma z} + (Z_L - Z_0)e^{\gamma z}}{(Z_L + Z_0)e^{-\gamma z} - (Z_L - Z_0)e^{\gamma z}}$$
$$= Z_0 \frac{Z_L(e^{\gamma z} + e^{-\gamma z}) - Z_0(e^{\gamma z} - e^{-\gamma z})}{Z_0(e^{\gamma z} + e^{-\gamma z}) - Z_L(e^{\gamma z} - e^{-\gamma z})}$$

Thus

$$Z_{in}(z) = Z_0 \frac{Z_L - Z_0 \tanh(\gamma z)}{Z_0 - Z_L \tanh(\gamma z)} \tag{2.30}$$

where

$$\tanh(\gamma z) = \frac{e^{\gamma z} - e^{-\gamma z}}{e^{\gamma z} + e^{-\gamma z}} \tag{2.31}$$

is the hyperbolic tangent function.

In practice, the input impedance is measured at a given distance l rather than at its z-axis value, as shown in Figure 2.5. Thus, the input impedance at l meters away from the load is

$$Z_{in}(l) = Z_0 \frac{Z_L + Z_0 \tanh(\gamma l)}{Z_0 + Z_L \tanh(\gamma l)} \tag{2.32}$$

Note that there is a sign change from Equation (2.30) since the distance should not be negative and we have used $l = -z$ and $\tanh(-\gamma l) = -\tanh(\gamma l)$.

If the loss of the transmission line can be neglected, that is $\gamma \approx j\beta$, Equation (2.32) can be simplified to

$$Z_{in}(l) = Z_0 \frac{Z_L + jZ_0 \tan(\beta l)}{Z_0 + jZ_L \tan(\beta l)} \tag{2.33}$$

This is a very useful equation. Special attention should be paid to the following cases:

- **Matched case:** $Z_L = Z_0$
  $Z_{in}(l) = Z_0$, the input impedance is the same as the characteristic impedance and is not a function of the length of the line.
- **Open circuit:** $Z_L = \infty$

$$Z_{in}(l) = Z_0 \frac{1}{j \tan(\beta l)} \tag{2.34}$$

The input impedance has no resistance, just reactance (capacitive for small l).

- **Short circuit:** $Z_L = 0$

$$Z_{in}(l) = jZ_0 \tan(\beta l) \tag{2.35}$$

Again, the input impedance has no resistance, just reactance (inductive for small l).

- **Quarter-wavelength case:** $l = \lambda/4$

$$Z_{in}(l) = \frac{Z_0^2}{Z_L} \tag{2.36}$$

This special case is called the *quarter-wavelength transform* since the load impedance is transformed (after a quarter wavelength) to the input impedence given by this simple equation. It is often used for impedance-matching purposes.

It should be pointed out that, in calculating the wavelength λ and wave number $\beta = 2\pi/\lambda$, the wavelength inside the transmission line is generally different from that in free space. The dielectric properties of the material of the transmission line have to be taken into account when doing such a calculation. The simplest case is that the wavelength is linked to the relative permittivity ε_r (also called the relative dielectric constant) of the material by

$$\lambda = \frac{\lambda_0}{\sqrt{\varepsilon_r}} \tag{2.37}$$

where λ_0 is the free space wavelength. More details will be given later in this chapter.

Example 2.1: Input impedance. A lossless transmission line with a characteristic impedance of 50 Ω is loaded by a 75 Ω resistor. Plot the input impedance as a function of the line length (up to two wavelengths).

Solution:

Since it is a lossless transmission line, Equation (2.33) is employed to calculate the input impedance. The result is shown in Figure 2.6 where both the resistance and reactance are plotted as a function of the normalized (to wavelength) line length. It is apparent that

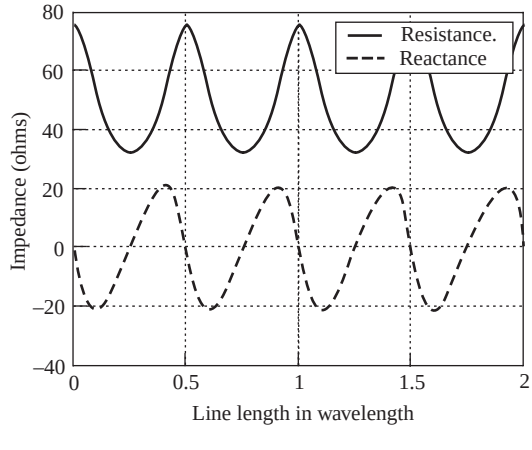


Figure 2.6 Input impedance as a function of the transmission line length for $Z_L = 75\ \Omega$ and $Z_0 = 50\ \Omega$

- The input impedance is a periodic function of the line length. The period is half of a wavelength.
- The input impedance is a complex number even when the load impedance is a pure resistance. The resistance changes between 75 Ω and 33 Ω, whilst the reactance changes between −20 Ω and +20 Ω.
- There are two resonant points (where the reactance is zero) over one period. These are at $l/\lambda = 0$ and 0.25 in this case.
- When $0 < l/\lambda < 0.25$, the reactance is negative, i.e. capacitive. When $0.25 < l/\lambda < 0.5$, the reactance is positive, i.e. inductive.

Example 2.2: Input impedance of a low-loss transmission line. A 75 Ω resistor is now connected to a good transmission line with characteristic impedance of 50 Ω. The attenuation constant is not zero but 0.2 Np/m at 1 GHz. Plot the input impedance as a function of the line length (up to 2λ). Assume that the effective relative permittivity is 1.5.

Solution:

For a low-loss transmission line, the characteristic impedance is still a constant (= 50 Ω in this case). The line length changes from 0 to 2λ. Since the effective relative permittivity is 1.5 and the frequency is 1 GHz, the wavelength in the medium is

$$\lambda = \frac{c}{f\sqrt{\varepsilon_r}} = \frac{3 \times 10^8}{10^9\sqrt{1.5}} \approx 0.245 \text{ (m)}$$

Thus, the length l is from 0 to 0.49 m.

The propagation constant is $\gamma = \alpha + j\beta = 0.2 + j2\pi/\lambda$. Using Equation (2.32), we can plot the input impedance, as shown in Figure 2.7. It can be seen that the only change is that the input impedance is no longer a periodic function of the line length. However, it still exhibits a period feature if we neglect the amplitude changes. All other features remain the same as in the lossless case.

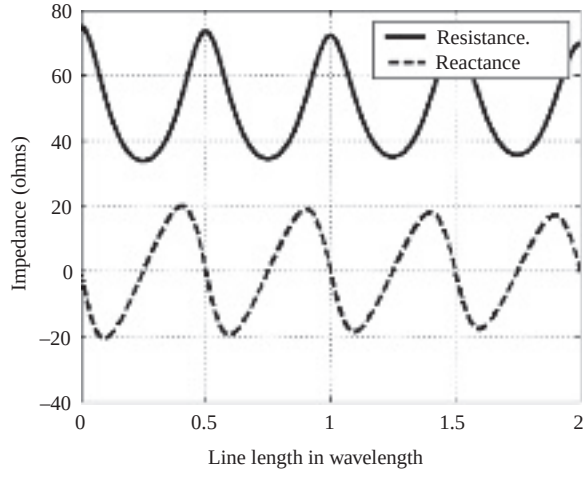


Figure 2.7 The input impedance along a low-loss transmission line for $Z_L = 75\ \Omega$ and $Z_0 = 50\ \Omega$

Example 2.3: Quarter-wavelength transform. A 75 Ω resistor is to be matched with a transmission line of characteristic impedance 50 Ω. If a quarter-wavelength transformer is employed, what should its characteristic impedance be?

Solution:
Using Equation (2.36), we have

$$Z_0 = \sqrt{Z_{in} Z_L} = \sqrt{75 \cdot 50} \approx 61.2\ (\Omega)$$

Since this is not a standard characteristic impedance of a transmission line, special dimensions and/or materials will be needed to construct this line.

2.2.3.2 The Reflection Coefficient and the Return Loss

The reflection coefficient was defined by Equation (2.29). If we replace the z-axis value by the length of the line l, it can be rewritten as

$$\Gamma(l) = \frac{Z_L - Z_0}{Z_L + Z_0} e^{-2\gamma l} = \Gamma_0 e^{-2\gamma l} \tag{2.38}$$

This is the voltage reflection coefficient. Since the power is proportional to the voltage squared, as shown in Equation (2.3), the power reflection coefficient is

$$\Gamma_P(l) = |\Gamma(l)|^2 = |\Gamma_0|^2 e^{-2\alpha l} \tag{2.39}$$

Both reflection coefficients are a good measure of how much signal/power is reflected back from the terminal. Obviously, when the load impedance is the same as the characteristic impedance, they are both zero and that is the matched case.

When the voltage reflection coefficient and power reflection coefficient are expressed in logarithmic forms, they give the same result, which is called the *return loss*:

$$L_{RT}(l) = -20 \log_{10}(|\Gamma(l)|) = -10 \log_{10}(\Gamma_P(l)) \tag{2.40}$$

Since the return loss should not be smaller than zero, there is a minus sign in Equation (2.40) (but in practice some people ignore the minus sign, which is not correct). A 3 dB return loss corresponds to the reflection coefficient being $1/\sqrt{2} \approx 0.707$ or the power reflection coefficient being 0.5.

It is worth noting the following special cases for a lossless transmission line:

- **Matched case:** $Z_L = Z_0$
  $\Gamma(l) = \Gamma_0 = 0$ – the reflection is zero at any point of the line (even for a non-perfect transmission line).
  $L_{RT} = \infty$ dB – in practice, this means that the return loss is huge, say 50 dB.

- **Open circuit:** $Z_L = \infty$
  $\Gamma_0 = 1$ and $\Gamma(l) = e^{-j2\beta l}$ – the amplitude of the reflection coefficient is 1 at any point of the line.
  $L_{RT} = 0$ dB – this means that all power is reflected back from the load.
- **Short circuit:** $Z_L = 0$
  $\Gamma_0 = -1$ and $\Gamma(l) = -e^{-j2\beta l}$ – the amplitude of the reflection coefficient is 1 at any reference point of the line; there is a phase shift of 180 degrees between the input and reflected voltages at the end of the line.
  $L_{RT} = 0$ dB – again, this means that all power is reflected back from the load.

Example 2.4: Reflection coefficient and return loss of a lossless transmission line. A 75 Ω resistor is connected to a lossless transmission line with characteristic impedance of 50 Ω.

a. What is the voltage reflection coefficient for $l = 0$ and $\lambda/4$, respectively?
b. What is the return loss for $l = 0$ and $\lambda/4$, respectively?

Solution:
For a lossless transmission line, the attenuation constant α is zero.

a. Using Equation (2.38), we have

$$\Gamma(0) = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{75 - 50}{75 + 50} = 0.2$$

and

$$\Gamma\left(\frac{\lambda}{4}\right) = \frac{Z_L - Z_0}{Z_L + Z_0} e^{-2\gamma l} = 0.2e^{-j2^*2\pi/4} = 0.2e^{-j\pi} = -0.2$$

This means that the phase of the reflection coefficient is changed by 180 degrees when the length of the transmission line is increased by a quarter-wavelength.
b. The return loss can be obtained by using Equation (2.40), i.e.

$$L_{RT} = -20\log_{10}(|\Gamma(l)|) = -10\log_{10}(\Gamma_P(l)) = 13.98 \ dB$$

for $l = 0$ and $\lambda/4$, and actually any length of such a transmission line.

Example 2.5: Reflection coefficient and return loss of a low-loss transmission line. A 75 Ω resistor is connected to a low-loss transmission line with characteristic impedance of 50 Ω. The attenuation constant is 0.2 Np/m at 1 GHz.

a. What is the voltage reflection coefficient for $l = 0$ and $\lambda/4$, respectively?
b. Plot the return loss as a function of the line length. Assume that the effective relative permittivity is 1.5.

Solution:

For this low-loss transmission line, the attenuation constant is $\alpha = 0.2$ Np/m. At 1 GHz, the wavelength in the line is

$$\lambda = \frac{c}{f\sqrt{\varepsilon_r}} = \frac{3 \times 10^8}{10^9\sqrt{1.5}} \approx 0.245 \text{ (m)}$$

a. Using Equation (2.38), we have

$$\Gamma(0) = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{75 - 50}{75 + 50} = 0.2$$

and

$$\Gamma\left(\frac{\lambda}{4}\right) = \frac{Z_L - Z_0}{Z_L + Z_0} e^{-2\gamma l} = 0.2e^{-2^*0.2^*0.245/4 - j2^*2\pi/4} = 0.1952e^{j\pi} = -0.1952$$

This means that not only the phase of the reflection coefficient but also the amplitude is changed when the length of the transmission line is increased by a quarter-wavelength.

b. The return loss can be obtained by using Equation (2.40), i.e.

$$L_{RT}(0) = -20\log_{10}(|\Gamma(l)|) = 13.98 \text{ dB}$$

and

$$L_{RT}\left(\frac{\lambda}{4}\right) = -20\log_{10}(|\Gamma(l)|) = 14.19 \text{ dB}$$

The return loss is slightly increased, as expected. Over two wavelengths, the change is shown in Figure 2.8. It follows $0.2e^{-2\alpha l}$ and is a straight line on a logarithmic scale, i.e. $L_{RT}(l) = 13.98 + 3.47l$ dB.

2.2.3.3 The Voltage Standing Wave Ratio (VSWR)

The VSWR (also known as the *standing wave ratio*, SWR) is defined as the ratio of the magnitude of the maximum voltage on the line to the magnitude of the minimum voltage on the line, as shown in Figure 2.9. Mathematically, it can be expressed as

$$VSWR(l) = \frac{|V|_{\max}}{|V|_{\min}} = \frac{|V_+| + |V_-|}{|V_+| - |V_-|} = \frac{1 + |\Gamma(l)|}{1 - |\Gamma(l)|} \tag{2.41}$$

Obviously, the VSWR is just another measure of how well a transmission line is matched with its load. Unlike the reflection coefficient, the VSWR is a scalar and has no phase information. For a nonperfect transmission line, the VSWR is a function of the length of the line (l) as well as the load impedance and the characteristic impedance of the line. But for a lossless transmission line, the VSWR is the same at any reference point of the line.

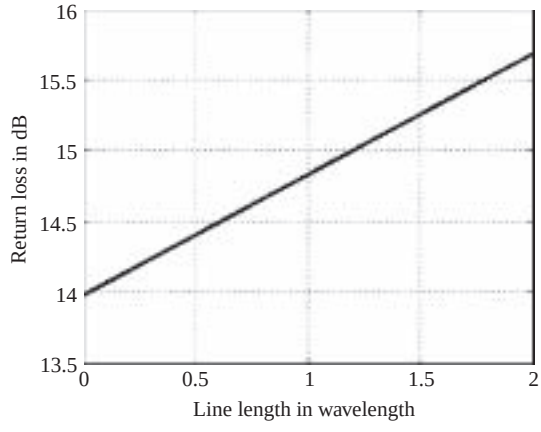


Figure 2.8 Return loss as a function of the line length

From Equation (2.41), we can prove that

$$|\Gamma| = \frac{VSWR - 1}{VSWR + 1} \tag{2.42}$$

This can be used to calculate the reflection coefficient once the VSWR is known.

Note the following special cases:

- **Matched termination** $Z_L = Z_o$:

$$VSWR = 1$$

- **Open circuit** $Z_L = \infty$:

$$VSWR = \infty$$

- **Short circuit** $Z_L = 0$:

$$VSWR = \infty$$

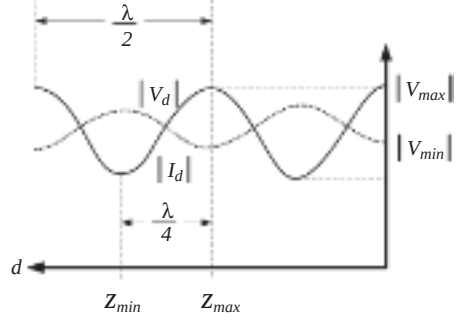


Figure 2.9 Standing waves of the voltage and current on a transmission line

Thus, the VSWR of a line is bounded by unity and infinity:

$$1 \leq VSWR \leq \infty$$

For most applications, the VSWR is required to be smaller than 2, which is considered a good match. But for the mobile phone industry, the desired VSWR is normally less than 3, which is due to the considerable effects of the human body on the performance of mobile phones.

Example 2.6: VSWR. A 75 Ω resistor is connected to a transmission line of characteristic impedance of 50 Ω. What is the VSWR at the termination?

Solution:
The reflection coefficient at the termination is

$$\Gamma(0) = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{75 - 50}{75 + 50} = 0.2$$

Using Equation (2.41) gives

$$VSWR(0) = \frac{1 + |\Gamma(0)|}{1 - |\Gamma(0)|} = 1.5$$

This is smaller than 2, thus this can be considered a very well-matched case.

In this section we have discussed the reflection coefficient, the return loss and VSWR. All these quantities are measures of impedance matching. Table 2.1 is a list of some typical values to show how they are interrelated. It is interesting to note that, if Z_L is a real number,

Table 2.1 Links between normalized impedance, reflection coefficient, return loss and VSWR

Z_L/Z_0	Γ	L_{RT}(dB)	*VSWR*	Note
∞	+1	0	∞	Open circuit
5.8470	0.7079	3	5.8470	Half power returned
3.0096	0.5012	6	3.0096	
1.9248	0.3162	10	1.9248	Close to $VSWR = 2$
1.2222	0.1000	20	1.2222	
1.0653	0.0316	30	1.0653	
1.0202	0.0100	40	1.0202	
1	**0**	∞	**1**	**Matched**
0.9802	−0.0100	40	1.0202	
0.9387	−0.0316	30	1.0653	
0.8182	−0.1000	20	1.2222	
0.5195	−0.3162	10	1.9248	Close to $VSWR = 2$
0.3323	−0.5012	6	3.0096	
0.1710	−0.7079	3	5.8470	Half power returned
0	−1	0	∞	Short circuit

- when $Z_L/Z_0 > 1$, $Z_L/Z_0 = VSWR$;
- when $Z_L/Z_0 < 1$, $Z_L/Z_0 = 1/VSWR$.

However, if Z_L is a complex number and its imaginary part is not zero, these simple relations linking Z_L/Z_0 and *VSWR* do not hold. For example, for $Z_L = j50$ ohms and $Z_0 = 50$ ohms, we have $\Gamma = j$, $L_{RT} = 0$ dB and $VSWR = \infty \neq Z_L/Z_0$.

2.3 The Smith Chart and Impedance Matching

2.3.1 The Smith Chart

The *Smith Chart*, as shown in Figure 2.10, was invented by Phillip H. Smith (1905–1987) and is a graphical aid designed for use by radio frequency (RF) engineers to solve transmission line and matching circuit problems. Although computer-aided tools have grown steadily over the years, the Smith Chart is still widely used today, not only as a problem-solving aid, but as a graphical demonstrator of how RF parameters behave and an alternative to using tabular information. The Smith Chart can be utilized to represent many parameters including impedances, admittances, reflection coefficients, scattering parameters, noise figure circles, constant gain contours and regions for unconditional stability. It is most frequently used at, or within, the unity radius region. However, the remainder is still mathematically relevant, being used, for example, in oscillator design and stability analysis.

The Smith Chart is plotted on the complex reflection coefficient plane in two dimensions, as shown in Figure 2.11. The horizontal axis is the real part of the reflection coefficient while the vertical axis shows the imaginary part of the reflection coefficient. The origin or center is

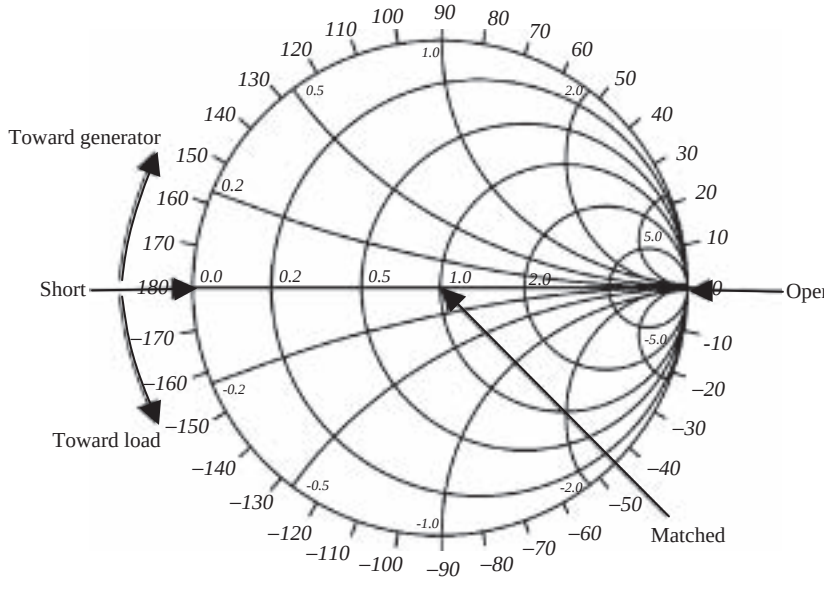


Figure 2.10 The standard Smith Chart

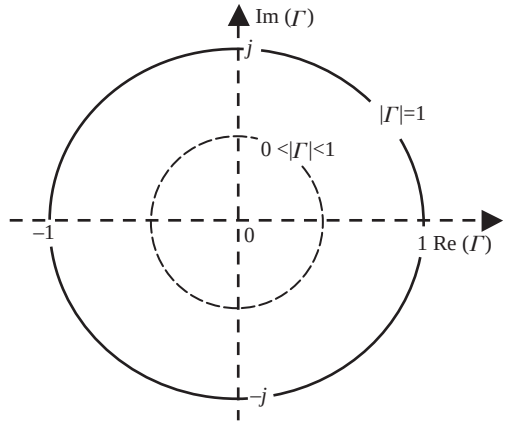


Figure 2.11 The Smith Chart showing the complex reflection coefficient

$|\Gamma| = 0$ (*VSWR* $= 1$). *In the standard Smith Chart, only the circle for* $|\Gamma| = 1$ *(VSWR* $= \infty$*) is shown and other circles are not displayed to make the chart simple and neat.*

Most information shown on the standard Smith Chart is actually the normalized complex impedance as shown in Figure 2.12: the resistance is displayed in circles and the reactance is shown as arched lines. The upper half space is inductive whilst the lower half space is capacitive. The middle line is pure resistance from 0 to infinity. If a reference point on a transmission line is moved away from the load (i.e. towards the source), this can be shown on the Smith Chart as the impedance point is moved on the $|\Gamma|$ (or VSWR) circle clockwise. When the reference point is moved towards the load, it means that the impedance point is moved on the $|\Gamma|$ circle anti-clockwise. The distance is normalized to the wavelength.

The Smith Chart can also be shown in other forms, such as normalized admittance, which is often known as the *Y Smith Chart*. In this book we are mainly interested in using the impedance Smith Chart. Other forms which can be found in references such as [1] are not discussed in this book.

As impedances change with frequency, problems using the Smith Chart can only be solved manually using one frequency at a time, the result being represented by a point. This is often adequate for narrowband applications (typically up to about 10% bandwidth) but for wide bandwidths it is usually necessary to apply Smith Chart techniques at more than one frequency across the operating frequency band. Provided that the frequencies are sufficiently close, the resulting Smith Chart points may be joined by straight lines to create a locus. A locus of points on a Smith Chart covering a range of frequencies can be employed to visually represent:

- how capacitive or inductive a load is across the frequency range;
- how difficult matching is likely to be at various frequencies;
- how well matched a particular component is.

The accuracy of the Smith Chart is reduced for problems involving a large spread of impedances, although the scaling can be magnified for individual areas to accommodate these.

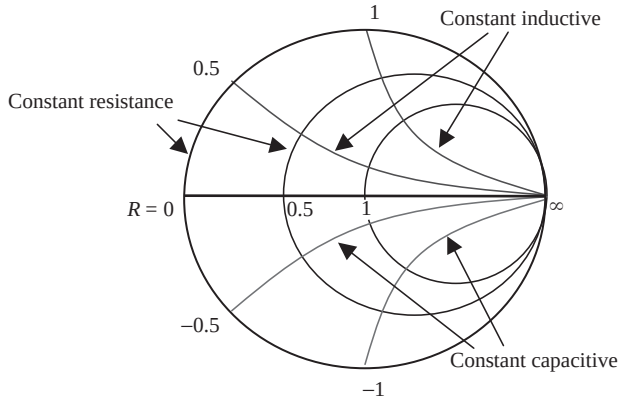


Figure 2.12 The Smith Chart showing the complex impedance

Example 2.7: Input impedance and reflection coefficient. Use a Smith Chart to redo Example 2.1, and also display the reflection coefficient on the chart.

Solution:

The characteristic impedance of the line is 50 Ω, thus the normalized load impedance is $75/50 = 1.5$, which can be uniquely identified as point A in the Smith Chart in Figure 2.13. The distance from the origin (center) to A is 0.2, thus the reflection coefficient $|\Gamma| = 0.2$. Moving the reference plane away from this load means moving point A along the $|\Gamma| = 0.2$ circle clockwise into the capacitive half space, and then past the resonant point at $R = 0.67$ (i.e. $0.67 \times 50\ \Omega = 33.5\ \Omega$) into the inductive half space. After 360 degrees (half wavelength), it is back to A. Over this period, the normalized reactance changes between – 0.4 and +0.4, i.e. $-20\ \Omega$ and $+20\ \Omega$. The results are the same as in Example 2.1.

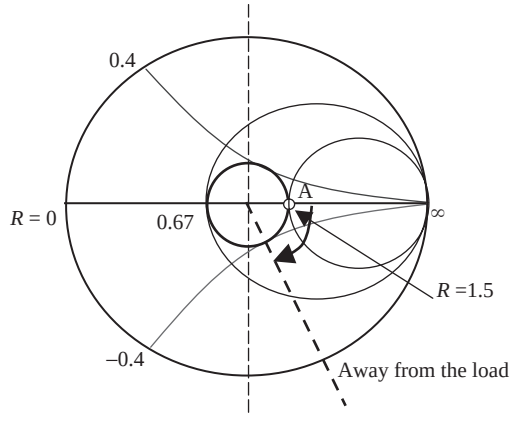


Figure 2.13 The Smith Chart showing the complex reflection coefficient

2.3.2 Impedance Matching

Impedance matching is the practice of making the output impedance of a source equal to the input impedance of the load in order to maximize the power transfer and minimize reflections from the load. Mathematically, it means the load impedance being the complex conjugate of the source impedance. That is,

$$Z_S = Z_L^* = (R_L + jX_L)^* = R_L - jX_L \tag{2.43}$$

When the imaginary part is zero, the two impedances are the same:

$$Z_S = Z_L^* = Z_L = R_L \tag{2.44}$$

Normally, we can use either lumped networks or distributed networks to match impedance.

2.3.2.1 Lumped Matching Networks

Lumped matching networks can be divided into three basic types: the L network, the T network and the pi (π) network. The aim of matching is to make the equivalent impedance of the network the same as the desired resistance R_{in} – usually the characteristic impedance of a transmission line. *Generally speaking, resistors are not employed for impedance matching* since the power could be consumed by the resistor without making a useful contribution to the operation of the system.

In Figure 2.14(a) an L network is employed to match the impedance R_{in}. It can be shown that this impedance is linked to the network elements by the following equations:

$$\begin{aligned} B &= \pm\frac{\sqrt{n-1}}{R_{in}} \\ X &= \pm\frac{R_{in}\sqrt{n-1}}{n} - X_L \end{aligned} \tag{2.45}$$

where $n = R_{in}/R_L$ and should be greater than 1. If this condition cannot be met, the second lumped L network, shown in Figure 2.14(b), may be used. The components are interrelated by

$$\begin{aligned} Y_L &= 1/Z_L = G_L + jB_L \\ B &= \pm\frac{\sqrt{m-1}}{mR_{in}} - B_L \\ X &= \pm R_{in}\sqrt{m-1} \end{aligned} \tag{2.46}$$

where $m = 1/(R_{in}G_L)$ and should be greater than 1.

It should be pointed out that lumped L networks have no degree of freedom to optimize the bandwidth, whilst the bandwidth is actually a very important consideration of many applications. To resolve this problem, one should consider adding further reactive elements to create T or π networks to optimize the overall bandwidth.

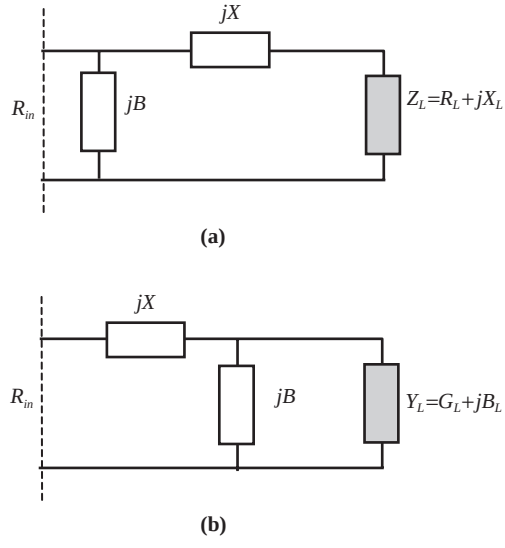


Figure 2.14 Lumped L networks (a) for $R_{in} > R_L$; (b) for $R_{in}G_L < 1$

Figure 2.15 is a T network, which may be viewed as another reactance (jX_2) added to the L network in Figure 2.14(a). The design process can be summarized by the following three steps:

Step 1: according to the load impedance and the desired bandwidth, choose X_1,

$$B_f = \frac{\Delta f}{f_o} = \frac{R_L}{|X_1 + X_L|} \tag{2.47}$$

Step 2: since Z_L and jX_1 are in series, the composite load impedance can be obtained as $Z_{LN} = Z_L + jX_1$.

Step 3: use Z_{LN} and the L network design in Equation (2.45) to find B and X_2.

Another option for impedance matching is to use a π network, as shown in Figure 2.16, which can be seen as an admittance (jB_2) added to the L network in Figure 2.14(b). The design process is very similar to that of the T network. That is:

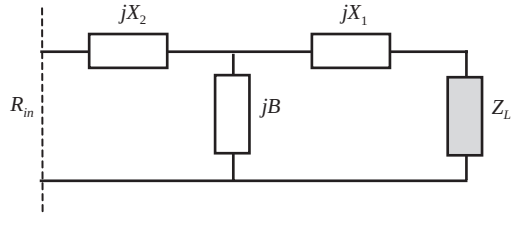


Figure 2.15 Lumped T network

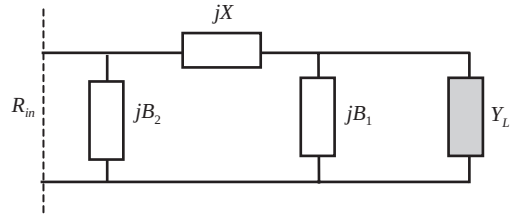


Figure 2.16 Lumped π network

Step 1: according to the load impedance and the desired bandwidth, choose B_1,

$$B_f = \frac{\Delta f}{f_o} = \frac{G_L}{|B_1 + B_L|} \tag{2.48}$$

Step 2: since Y_L and jB_1 are in parallel, the composite load admittance can be obtained as $Y_{LN} = Y_L + jB_1$.

Step 3: use Y_{LN} and the L network design in Equation (2.46) to find X and B_2.

Example 2.8: Impedance matching. A load with an impedance of $10 - j100\ \Omega$ is to be matched with a 50 Ω transmission line. Design a matching network and discuss if there are other solutions available.

Solution:

Since $Z_L = R_L + jX_L = 10 - j100$ and $n = R_{in}/R_L = 50/10 = 5 > 1$, the L network in Figure 2.14(a) is a suitable matching network. Using Equation (2.45) we obtain

$$B = \pm\frac{\sqrt{n-1}}{R_{in}} = \pm\frac{2}{50} = \pm 0.04$$

$$X = \pm\frac{R_{in}\sqrt{n-1}}{n} - X_L = \pm\frac{50 \times 2}{5} + 100 = \pm 20 + 100$$

Thus, there are two sets of solutions: $(B,X) = (0.04, 120)$ and $(B,X) = (-0.04, 80)$.

Now let us see if there are other possible solutions. Because $Y_L = 1/(R_L + jX_L) \approx 0.001 + j0.001$ and $m = 1/(R_{in}G_L) = 1/0.05 = 20 > 1$, we can also use the L network in Figure 2.14(b) to match the impedance:

$$B = \pm\frac{\sqrt{m-1}}{mR_{in}} - B_L = \pm\frac{\sqrt{19}}{20 \times 50} - 0.001 \approx \pm 0.0043 - 0.001$$

$$X = \pm R_{in}\sqrt{m-1} = \pm 50\sqrt{19} \approx \pm 217.9$$

Therefore, we have obtained another two sets of solutions: $(B,X) = (0.0033, 217.9)$ and $(B,X) = (-0.0053, -217.9)$.

In addition, we can also use T and π networks to match the load with the transmission line and the bandwidth can be controlled.

It is apparent that, for a given load impedance, there is more than one matching network available. In practice, the decision as to which network to adopt normally depends on some other parameters such as the bandwidth (to be discussed in Example 2.9) and values of the elements.

2.3.2.2 Distributed Matching Networks

Distributed matching networks can be formed by a quarter-wavelength transmission line, an open-circuit transmission line, a short-circuit transmission line or their combinations. They can be represented mathematically by Equations (2.34) to (2.36). The process is best visualized on the Smith Chart.

The quarter-wavelength transformer mentioned earlier is a unique and popular narrowband impedance-matching technique and the process is quite straightforward. The short- and open-circuit stub tuning is very similar to the lumped matching circuit tuning: reactance and/or admittance is added to the matching network. The example below is a good illustration of how to realize matching using a distributed network.

Example 2.9: Impedance matching and bandwidth. A load with an impedance of $10 - j100\ \Omega$ is to be matched with a $50\ \Omega$ transmission line. Design two distributed matching networks and compare them in terms of the bandwidth performance.

Solution:

The normalized load impedance is:

$$z_L = Z_L/50 = 0.2 - j2$$

As shown in Figure 2.17, this corresponds to a unique point A on the Smith Chart. The reflection coefficient is

$$\Gamma = \frac{Z_L - 50}{Z_L + 50} = 0.5588 - j0.7353$$

$$|\Gamma| = 0.9235$$

The $|\Gamma| = 0.9235$ circle is shown in Figure 2.17 as a broken line. To match the impedance, we can move this point (A) clockwise (towards the source) along this circle to points B1, B2, B3 or B4, which are the crossover points with the circle of $z = 1$ or $1/z = 1$. It can then be further moved either along circle $z = 1$ or $1/z = 1$ to the center O–the matching point. The normalized admittances at B1 and B2 are -0.04 and $+0.04$, and the normalized reactances at B3 and B4 are $+217.9$ and -217.9, respectively. These four routes correspond to the four solutions from the L networks, as demonstrated in Example 2.8.

The matching network using B1 or B2 can be illustrated by Figure 2.18(a), where the open- or short-circuit stub is in parallel with the load impedance, while the matching network using B3 or B4 can be shown by Figure 2.18(b), where the open- or short-circuit stub is in series with the load impedance. The rotational angles (from A to B1, B2, B3 or B4) on the Smith Chart determine the stub lengths l_1 and l_2. There are at least four possible designs.

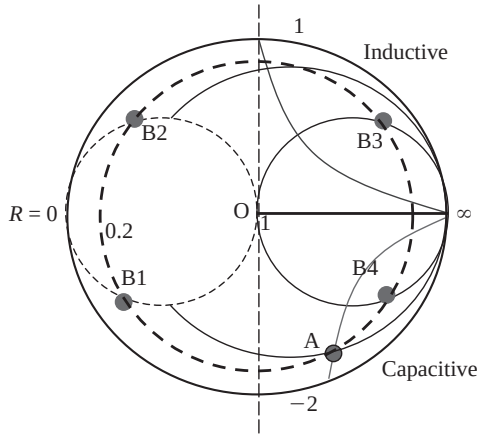


Figure 2.17 Impedance matching using a Smith Chart

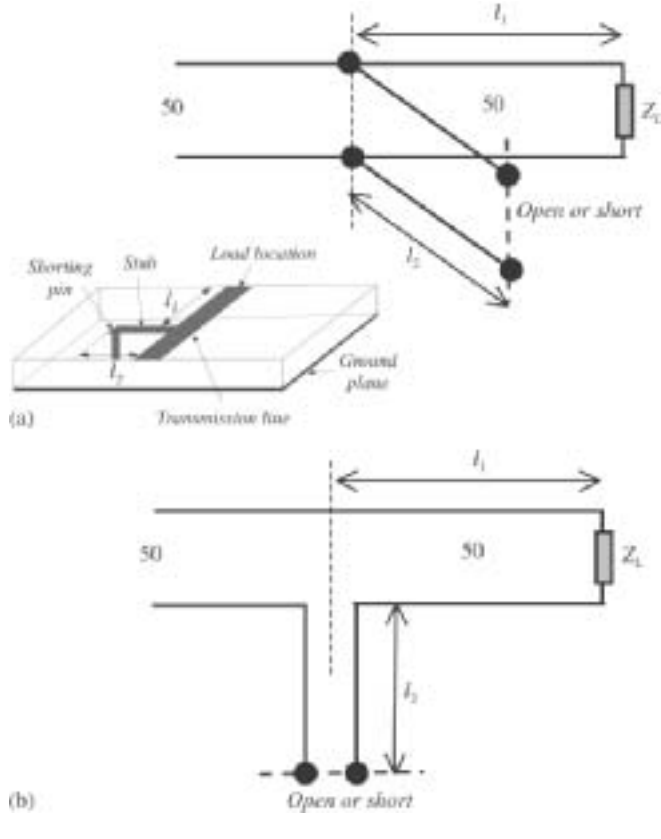


Figure 2.18 Stub-matching networks (a) parallel stub matching; (b) series stub matching

To design a matching network using B1 we can follow the steps below:

Step 1: Move from point A to B1, the rotational angle is about 0.582π (i.e. 104.8°) on the Smith Chart, which corresponds to $l_1 = 0.1455\lambda$. Note that the rotational angle is twice βl_1 ($= 0.291\pi$, not 0.582π, *since the impedance period is* π, *not* 2π). The impedance at point B1 is $0.0413 - j0.1984$ while its admittance is $1.0 + j4.8318$.

Step 2: Move from point B1 to the center O. This can be achieved easily using a stub connected in parallel with the line, thus it is advantageous to work in admittances. The stub in parallel with the line should produce a susceptance of -4.8318. This can be achieved by the following two designs:

A. a short circuit with a stub length $l_2 = 0.0325\lambda$;
B. an open circuit with a stub length $l_2 = 0.2825\lambda$.

The stub length can also be obtained from the Smith Chart.

Now let us examine the bandwidth of these two designs. Assuming the center frequency is 1 GHz, we have $l_1 = 0.1455\lambda = 4.365$ cm, and $l_2 = 0.0325\lambda = 0.975$ cm for Design A, and $l_2 = 0.2825\lambda = 8.475$ cm for Design B. We can use the input impedance formula:

$$Z_{in}(l) = Z_0 \frac{Z_L + jZ_0 \tan(\beta l)}{Z_0 + jZ_L \tan(\beta l)}$$

and the following two equations:

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}$$

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|}$$

to obtain the VSWR as a function of the frequency. The results are shown in Figure 2.19. It is apparent that

- both designs have an excellent impedance match at the center frequency 1 GHz;
- the stub length of Design A is shorter than that of Design B whilst the bandwidth of Design A is much wider than that of Design B. This is a very interesting and useful result.

Similar conclusions can be drawn from the matching networks built at the other points (B2, B3 and B4). To build a distributed matching network, the length should be as short as possible to maximize the bandwidth.

The frequency bandwidth limitation on matching networks has been investigated by many people. There exists a general limit on the bandwidth over which an arbitrarily good impedance match can be obtained in the case of a complex load impedance. It is related to the ratio of reactance to resistance, and to the bandwidth over which we wish to match the load.

Figure 2.20 shows four load impedances (series RL, series RC, parallel RC and parallel RL) with matching networks, which are specific examples of the L matching network discussed earlier. Take the parallel RC load impedance as an example; Bode and Fano derived, for lumped

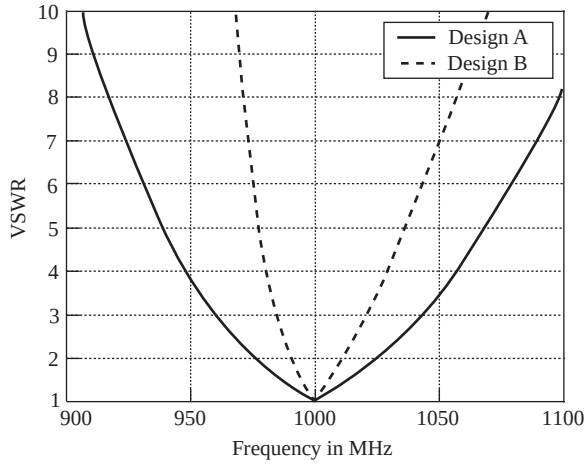


Figure 2.19 VSWR of different designs as a function of frequency

circuits, a fundamental limitation for this, which can be expressed as [2]

$$\int_0^\infty \ln\left(\frac{1}{|\Gamma(\omega)|}\right) d\omega \leq \frac{\pi}{RC} \tag{2.49}$$

This is known as the *Bode–Fano limit* for parallel RC. Since $\ln(1) = 0$, there is no contribution to this integral over frequencies for $|\Gamma| = 1$, so it can be seen that it is desirable to have the maximum mismatch outside the band of interest if a broad bandwidth is required. If this condition is assumed, the integral is limited to the bandwidth of interest ($\Delta\omega$), and we can get an idea of how well we can match an arbitrary complex impedance over that bandwidth. For an idealized case, this equation can be simplified to

$$\Delta\omega \ln\left(\frac{1}{|\Gamma(\omega)|}\right) \leq \frac{\pi}{RC} \tag{2.50}$$

This clearly shows how the bandwidth ($\Delta\omega$) is linked to the matching (Γ) and load impedance (RC). For a given reflection coefficient (or VSWR) and the RC product, one can estimate the maximum bandwidth attainable.

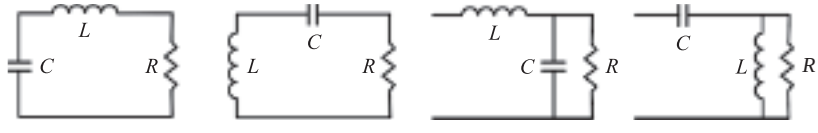


Figure 2.20 Four load impedances with LC matching networks

Similarly, the limits for the other three matching LC networks were obtained as:

Series RL: $\Delta\omega \ln(\frac{1}{|\Gamma(\omega)|}) \leq \frac{\pi}{R/L}$
Series RC: $\Delta\omega \ln(\frac{1}{|\Gamma(\omega)|}) \leq \pi\omega_0^2 RC$
Parallel RL: $\Delta\omega \ln(\frac{1}{|\Gamma(\omega)|}) \leq \frac{\pi\omega_0^2}{L/R}$

There are similar limitations on other forms of complex impedance. A general implication of the Bode–Fano limit is that one should not waste any match out-of-band, and that the best in-band match is obtained with Tchebyscheff rather than maximally flat networks. The best broadband impedance-matching practice incorporates the complex load impedance into, for example, a multisection filter structure with a design that includes the characteristics of the load.

2.3.3 The Quality Factor and Bandwidth

The bandwidth is indeed a very important parameter for any electric/electronic component and system. It is closely linked to the *quality factor, Q*, which is a measure of how much lossless reactive energy is stored in a circuit compared to the average power dissipated.

Antenna bandwidth is maximized when the power dissipation is comparatively high. In other words, a low Q is required for wide bandwidths. In turn, the extent to which this energy is associated with radiation (rather than conductor or dielectric losses) determines the antenna efficiency. For a circuit component such as an inductor or capacitor, we require the resistive losses to be low; hence Q is required to be high. It is often the case that antennas and circuit components have seemingly contradictory requirements: *antennas are designed to have a low Q, whereas circuit components are designed for a high Q*.

The quality factor is quoted as being either unloaded or loaded. For the latter, the losses of the external circuit – for example, the source – are included, whereas for the former they are not. The unloaded quality factor, Q, is defined as

$$Q \equiv \omega \frac{\text{(total energy stored)}}{\text{(average power loss in the load)}} = \omega \frac{W_E + W_M}{P_L} \tag{2.51}$$

where W_E is the energy stored in the electric field, W_M is the energy stored in the magnetic field and P_L is the average power delivered to the load. The loaded quality factor, Q_L, can also be given by Equation (2.51) but with P_L replaced by the total power P_T, which is dissipated in both the external circuit and the load.

At resonance, the electric and magnetic field energies have the same magnitudes and the formulas simplify such that the unloaded quality factor at resonance Q_0 is given by

$$Q_0 \equiv \frac{2\omega_0 W_E}{P_L} = \frac{2\omega_0 W_M}{P_L} \tag{2.52}$$

where ω_0 is the angular resonant frequency ($= 2\pi f_0$, where f_0 is the resonant frequency).

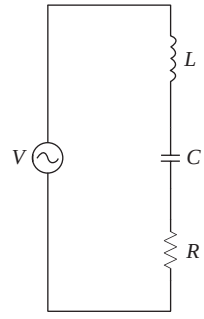


Figure 2.21 Series resonant circuit. (Reproduced by permission of Delft University Press)

There are many simplifications of these definitions that are widely used, but which only truly apply when they are derived directly from the above. For example, a relation that is often used is

$$Q_0 = \frac{f_0}{f_2 - f_1} = \frac{1}{B_F} \tag{2.53}$$

where f_1 and f_2 are the frequencies at which the power reduces to half of its maximum value at the resonant frequency, f_0, and where B_F is the *fractional bandwidth*. This relation only truly applies to simple circuits, but is considered to be a good approximation to a wide range of unloaded configurations. It has the advantage of providing a simple relationship between Q and fractional bandwidth, but we will see later that it only accurately applies to simple, single resonant circuits and should therefore be used with some care. The derivation of this relation is illustrated below.

Consider the series resonant circuit shown in Figure 2.21. The power dissipated in the resistance R is proportional to the square of the magnitude of the current I. This is plotted in Figure 2.22 with $R = 50\ \Omega$, $L = 79.5775$ nH and $C = 0.3183$ pF. The half-power frequencies,

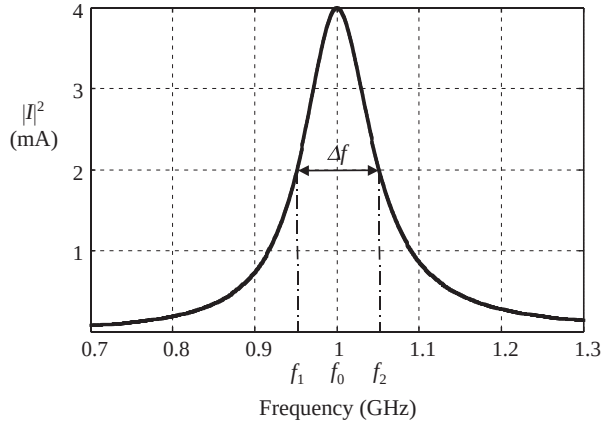


Figure 2.22 Relative power dissipated in a series resonant circuit around resonance

f_1 and f_2, are found to be 0.9513 GHz and 1.0513 GHz respectively. As shown in Figure 2.22, the resonant frequency (when the current is maximized) is 1 GHz. Using Equation (2.53) yields a Q_0 of 10.

It is possible to find the unloaded quality factor of the circuit directly from the stored energies of the inductor and capacitor. The magnetic and electric energies are given by

$$W_M = \frac{1}{2} L I^2,\ W_E = \frac{1}{2} C V_C^2 \tag{2.54}$$

where V_C is the voltage across the capacitor. Writing these and the power delivered to the load (i.e. the resistor, R) in terms of the current I gives

$$W_M = \frac{1}{2} L I^2, \quad W_E = \frac{1}{2} C \frac{1}{(\omega C)^2} I^2, \quad P_L = R I^2 \tag{2.55}$$

Further substitution in Equation (2.51) yields

$$Q = \frac{\left(\omega L + {}^1\!/_{\omega C}\right)}{2R} \tag{2.56}$$

At resonance, the magnitudes of the electric and magnetic energies are equal and this relation simplifies to

$$Q_0 = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 C R},\ \text{and}\ \omega_0 = \frac{1}{\sqrt{LC}} \tag{2.57}$$

This formula is often used, but it should be understood that it only applies at resonance. Taking $R = 50\ \Omega$, $L = 79.5775$ nH and $C = 0.3183$ pF gives $Q_0 = 10$, as previously.

It is interesting to evaluate the ratio of Q at any frequency to that at resonance. From Equations (2.56) and (2.57), this is given by

$$q = \frac{Q}{Q_0} = \frac{1}{2}\left(\frac{\omega}{\omega_0} + \frac{\omega_0}{\omega}\right) \tag{2.58}$$

It can be seen that the unloaded quality factor is a minimum at resonance, although the variation with frequency is slow. For bandwidths of less than 20%, Q is approximately equal to Q_0 with an error of less than 0.5% – i.e. the term in the parentheses of Equation (2.58) is approximately equal to two over moderate bandwidths.

We would like to derive a relationship between the unloaded quality factor and the bandwidth of the circuit. The bandwidth is normally taken to be the range of frequencies over which the power dissipated; P_L is greater than half of the maximum, P_{L0} (at resonance). However, in the sections that follow more general relations are derived based on a specified power transfer to the load.

The current in the circuit is given by

$$I = \frac{V}{R + j\left(\omega L - {}^{1}\!/_{\omega C}\right)} \tag{2.59}$$

From Equation (2.59), the ratio of the power dissipated at any frequency to the power dissipated at resonance is given by

$$p = \frac{P_L}{P_{L0}} = \left|\frac{I}{I_0}\right|^2 = \frac{1}{1 + \left(\dfrac{\omega L}{R} - \dfrac{1}{\omega CR}\right)^2} \tag{2.60}$$

This can be written

$$p = \frac{1}{1 + \chi^2} \tag{2.61}$$

where, from (2.57),

$$\chi = Q_0\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \tag{2.62}$$

The relation given in Equation (2.61) can be solved to give

$$Q_0\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) = \pm\sqrt{\frac{1-p}{p}} \tag{2.63}$$

This is a quadratic equation in ω with two positive and two negative solutions. The difference between the two positive solutions is

$$\omega_2 - \omega_1 = \frac{\omega_0}{Q_0}\sqrt{\frac{1-p}{p}} \tag{2.64}$$

This gives the fractional bandwidth as

$$B_F = \frac{f_2 - f_1}{f_0} = \frac{1}{Q_0}\sqrt{\frac{1-p}{p}} \tag{2.65}$$

When $p = 0.5$, this simplifies to the familiar expression

$$B_F = \frac{1}{Q_0} \tag{2.66}$$

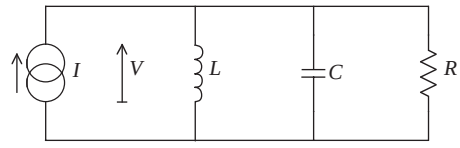


Figure 2.23 Parallel anti-resonant circuit. (Reproduced by permission of Delft University Press)

as previously seen in Equation (2.53). This derivation does not make any assumptions about the numerical value of the quality factor or the bandwidth and is, therefore, applicable to both high- and low-Q systems.

The response of a parallel resonant circuit – shown in Figure 2.23 – can be found in much the same way as for series resonance.

The energies stored in the electric and magnetic fields of the capacitor and inductor respectively and the power dissipated in the resistance are given by

$$W_E = \frac{1}{2} C V^2, \quad W_M = \frac{1}{2} L \frac{1}{(\omega L)^2} V^2, \quad P = G V^2 \tag{2.67}$$

where G is the conductance. Substitution in Equation (2.51) gives

$$Q = \frac{\left(\omega C + {}^{1}\!/_{\omega L}\right)}{2G}. \tag{2.68}$$

At anti-resonance, the magnitudes of the electric and magnetic energies are equal and this relation simplifies to

$$Q_0 = \frac{\omega_0 C}{G} = \frac{1}{\omega_0 L G} \tag{2.69}$$

The ratio between Q and Q_0 is the same as for the series resonance, given by Equation (2.58). It can also be shown (using the same method) that the fractional bandwidth is the same as for the series resonant circuit.

The formulas for Q that have been derived so far have been for series and parallel resonant circuits respectively. However, they are often applied – strictly incorrectly – to other circuit combinations. We will see later that this can give large errors, so some care is required in applying these formulas.

2.4 Various Transmission Lines

There are many transmission lines developed for various applications. The most popular ones are shown in Figure 2.24. They are the two-wire transmission line, the coaxial cable, the microstrip, the stripline, the coplanar waveguide (CPW) and the waveguide. We are going to examine these transmission lines in terms of their characteristic impedance, basic mode, frequency bandwidth, loss characteristic and costs.

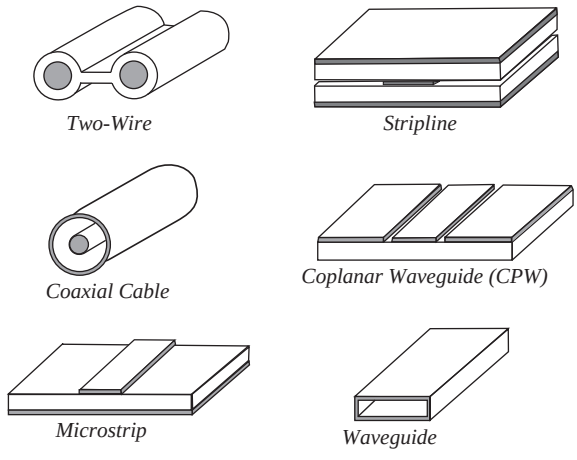


Figure 2.24 Various transmission lines

2.4.1 Two-wire Transmission Line

This is the simplest transmission line and its cross-sectional view is given in Figure 2.25. The separation of the wires is D and the diameter of the wires is d. The medium between the wires has a permittivity of ε.

It can be shown that the per unit length inductance and capacitance of the transmission line are [2]:

$$L = \frac{\mu}{\pi} \ln \frac{D + \sqrt{D^2 - d^2}}{d}, \qquad C = \frac{\pi \varepsilon}{\ln \dfrac{D + \sqrt{D^2 - d^2}}{d}} \tag{2.70}$$

If the medium between the wires has a conductivity of σ_1, and the conductivity of the wire is σ_2, we can obtain the resistance and conductance of a unit length line as

$$R = \frac{2}{\pi d} \sqrt{\frac{\omega \mu}{2\sigma_2}}, \quad G = \frac{\pi \sigma_1}{\ln \dfrac{D + \sqrt{D^2 - d^2}}{d}} \tag{2.71}$$

respectively.

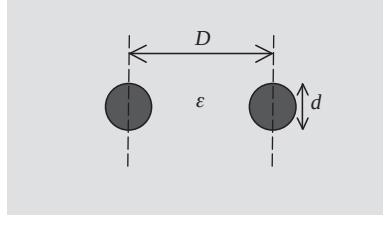


Figure 2.25 Two-wire transmission line

2.4.1.1 Characteristic Impedance

If the loss of the line can be considered very small, the characteristic impedance is given by Equation (2.22), i.e.

$$\begin{aligned} Z_0 &= \sqrt{\frac{L}{C}} = \sqrt{\frac{\mu}{\pi^2 \varepsilon}} \ln \frac{D + \sqrt{D^2 - d^2}}{d} \\ &\approx \frac{120}{\sqrt{\varepsilon_r}} \ln \frac{D + \sqrt{D^2 - d^2}}{d} \end{aligned} \tag{2.72}$$

The typical value of industrial standard lines is 300 Ω. This type of transmission line was commonly used to connect a television receiving antenna (usually a folded dipole with an impedance around 280 Ω) to a home television set many years ago. The uniform spacing is assured by embedding the two wires in a low-loss dielectric, usually polyethylene. Since the wires are embedded in the thin ribbon of polyethylene, the dielectric space is partly air and partly polyethylene.

2.4.1.2 Fundamental Mode

The electromagnetic field distribution around the two-wire transmission line is illustrated by Figure 1.15. Both the electric field and magnetic field are within the transverse (to the propagation direction) plane, thus this mode is called the *TEM* (transverse electro magnetic) *mode*. This means that it is nondispersive and the velocity is not changed with the frequency. The plane wave can also be considered a TEM wave.

2.4.1.3 Loss

Since the lumped parameters of a transmission line are given by Equations (2.70) and (2.71), the attenuation constant α can be calculated using Equation (2.14). However, the principal loss of the two-wire transmission line is actually due to radiation, especially at higher frequencies. Thus, this type of transmission line is not suitable for higher frequency applications. The typical usable frequency is less than 300 MHz.

Some people may be familiar with the *twisted-pair transmission line*. As the name implies, the line consists of two insulated wires twisted together to form a flexible line without the use of spacers. It has relatively good EMC (electromagnetic compatibility) performance – the twisted configuration cancels out the radiation from both wires and results in a small and symmetrical total field around the line; but it is not suitable for high frequencies because of the high dielectric losses that occur in the rubber insulation (low costs) as well as the radiation. When the line is wet, the losses increase significantly.

2.4.2 Coaxial Cable

The coaxial cable consists of a central, insulated wire (inner conductor) mounted inside a tubular outer conductor, as shown in Figure 2.26. In some applications, the inner conductor is also tubular. The inner conductor is insulated from the outer conductor by insulating materials

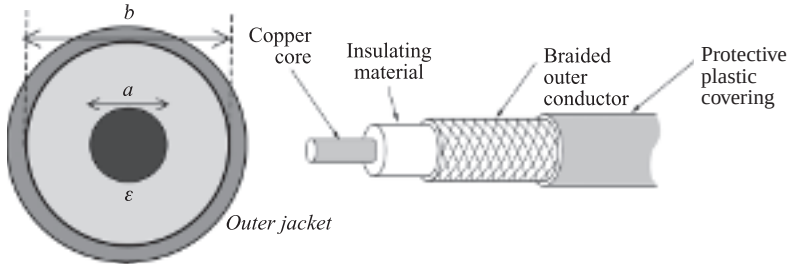


Figure 2.26 The configuration of a coaxial line

which are made of Pyrex, polystyrene, polyethylene plastic or some other material that has good insulating characteristics and low dielectric losses at high frequencies over a wide range of temperatures. In order to ensure good EMC performance, shielded and double-shielded coaxial cables have been developed and are available on the market. This type of transmission line is widely used for RF engineering and antenna measurements and for the connection between the antenna and transceiver.

Coaxial cables come in three basic types: flexible, semi-rigid and rigid. The rigid cable gives the best performance and is normally for high-performance and phase-sensitive applications, whilst the flexible cables are cheap and, obviously, flexible. The semi-rigid cable is a compromise. As shown in Figure 2.26, the diameters of the inner and outer conductors of a cable are denoted a and b respectively, and the relative permittivity of the insulating material is ε_r. The dielectric material reduces the velocity of the wave inside the cable to $c/\sqrt{\varepsilon_r}$. Some common loading materials and corresponding velocities are shown in Table 2.2 [3].

If the conductivities of the insulating material are zero and it can be shown that the per unit length parameters of the coaxial line are

$$\begin{aligned} L &= \frac{\mu}{2\pi}\ln\frac{b}{a} \qquad & R &= \sqrt{\frac{f\mu}{\pi\sigma_2}}\left(\frac{1}{a}+\frac{1}{b}\right) \\ C &= 2\pi\varepsilon \Big/ \ln\frac{b}{a} & G &= 2\pi\sigma_1 \Big/ \ln\frac{b}{a} \end{aligned} \tag{2.73}$$

Table 2.2 Coaxial cable material and velocity

Dielectric type	Time delay (ns/m)	Propagation velocity (% of c)
Solid polyethylene (PE)	5.05	65.9
Foam polyethylene (FE)	4.17	80.0
Foam polystyrene (FS)	3.67	91.0
Air space polyethylene (ASP)	3.77–3.97	84–88
Solid Teflon (ST)	4.79	69.4
Air space Teflon (AST)	3.71–3.94	85–90

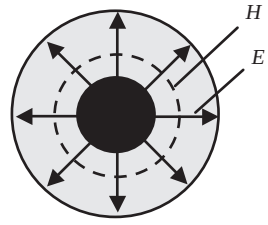


Figure 2.27 Field distribution within a coaxial line

The velocity, as shown in Equation (2.25), is

$$v = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c}{\sqrt{\varepsilon_r}}$$

2.4.2.1 Characteristic Impedance

Normally, if the loss of the line can be considered very small, the characteristic impedance is given by Equation (2.22), i.e.

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{\sqrt{\mu/\varepsilon}}{2\pi} \ln\frac{b}{a} = \frac{60}{\sqrt{\varepsilon_r}} \ln\frac{b}{a} \tag{2.74}$$

The typical value for industrial standard lines is 50 Ω or 75 Ω.

2.4.2.2 Fundamental Mode

The electromagnetic field distribution around the coaxial cable is illustrated by Figure 2.27. Again, both the electric field and magnetic field are within the transverse (to the propagation direction) plane, thus this field is *TEM mode*.

However, this is only true for the frequencies below the *cut-off frequency*, which is [2]

$$f_c = \frac{v}{\pi(a+b)} \tag{2.75}$$

where v is the velocity of the wave in the cable and the cut-off wavelength is

$$\lambda_c = \pi(a+b) \tag{2.76}$$

If the operational frequency is above this cut-off frequency (or, say, the operational wavelength is below the cut-off wavelength), the field within the coaxial cable may no longer be TEM mode; some higher modes such as TE_{11} (a transverse electric field, the magnetic field has nontransverse component) mode may exist, which is not a desirable situation since the loss could be significantly increased.

2.4.2.3 Loss

Since the transmission line lumped parameters are given by Equation (2.73), the attenuation constant α can be calculated using Equation (2.14) or Equation (2.23) for a low-loss line; that is

$$\alpha \approx \frac{R}{2Z_0} + \frac{GZ_0}{2} = \frac{\sqrt{f\mu\varepsilon_r}(1/a + 1/b)}{120\sqrt{\pi\sigma_2}\ln(b/a)} + \frac{60\pi\sigma_1}{\sqrt{\varepsilon_r}} \tag{2.77}$$

which is a function of the coaxial dimensions a and b as well as the conductivity and permittivity of the materials. When $b/a \approx 3.592$ (which means that the typical characteristic impedance should be around 77 ohms), the attenuation reaches the minimum. This is one of the most important considerations when making the cable.

In addition to the characteristic impedance, mode and loss of a cable, there are some other considerations when choosing a transmission line. Power-handling capacity is one of them, since it is very important for radar and high-power applications. The breakdown electric field strength in air is about 30 kV/cm (this means the best characteristic impedance should be close to 30 ohms). A list of some commercial cables with some important specifications is given in Table 2.3, where OD stands for the outer diameter and Vmax is the maximum voltage which may be applied to the cable in Volts. A more complete list can be found in [3]. It is clear that there are over 100 industry standard cables on the market with various specifications. The one with the smallest loss at 400 MHz is RG-211A, about 2.3 dB/100 ft. It is also one of the most expensive cables – normally the cost is inversely proportional to the loss of the cable.

2.4.3 Microstrip Line

As shown in Figure 2.28, a microstrip line may be viewed as a derivative of a two-wire transmission line and is perhaps the most widely used form of planar transmission line. One side of the structure is freely accessible for the mounting of packaged devices and the geometry lends itself extremely well to PCB patterning techniques to define the circuit. It has been used extensively in microwave and millimeter circuits and systems.

Due to the complexity of the structure, the analytical expressions of per unit length parameters are difficult to obtain. The effective relative permittivity is approximated as

$$\varepsilon_{re} \approx \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2\sqrt{1 + 12d/W}} \tag{2.78}$$

This is an empirical expression and is a function of the material property and the ratio W/d. W is the width of the strip and d is the thickness of the substrate, which has a relative permittivity ε_r.

2.4.3.1 Characteristic Impedance

The calculation of the characteristic impedance is not an easy task. From the transmission line theory, the relation between the velocity and per unit length inductance and capacitance is

$$v = \frac{1}{\sqrt{LC}} = \frac{c}{\sqrt{\varepsilon_{re}}}$$

Table 2.3 Some commercial coaxial cables and their specifications (Reproduced by permission of RFCafe.com)

Type (/U)	MIL-W-17	Z_0 (Ω)	Dielectric type	Capacitance (pF/ft)	OD (in.)	dB/100 ft @400 MHz	Vmax (rms)	Shield
RG-6A	/2-RG6	75.0	PE	20.6	0.332	6.5	2700	Braid
RG-8		52.0	PE	29.6	0.405	6.0	4000	Braid
RG-8A		52.0	PE	29.6	0.405	6.0	5000	Braid
RG-9		51.0	PE	30.2	0.420	5.9	4000	Braid
RG-11A	/6-RG11	75.0	PE	20.6	0.405	5.2	5000	Braid
RG-55B		53.5	PE	28.8	0.200	11.7	1900	Braid
RG-58A	/28-RG58	52.0	PE	29.6	0.195	13.2	1900	Braid
RG-58C	/28-RG58	50.0	PE	30.8	0.195	14.0	1900	Braid
RG-59/A	/29-RG59	73.0	PE	21.1	0.242	10.5	2300	Braid
RG-59B	/29-RG59	75.0	PE	20.6	0.242	9.0	2300	Braid
RG-141/A		50.0	ST	29.4	0.190	9.0	1900	Braid
RG-142/A/B	/60-RG142	50.0	ST	29.4	0.195	9.0	1900	Braid
RG-164	/64-RG164	75.0	PE	20.6	0.870	2.8	10,000	Braid
RG-174		50.0	ST		0.100	17.3	1200	Braid
RG-177	/67-RG177	50.0	PE	30.8	0.895	2.8	11,000	Braid
RG-178/A/B	/93-RG178	50.0	ST	29.4	0.072	29.0	1000	Braid
RG-180A/B	/95-RG180	95.0	ST	15.4	0.140	17.0	1500	Braid
RG-188		50.0	ST		0.050	17.5	700	Braid
RG-211/A	/72-RG211	50.0	ST	29.4	0.730	2.3	7000	Braid
RG-223	/84-RG223	50.0	PE	19.8	0.211	8.8	1900	Dbl Braid
RG-316	/113-RG316	50.0	ST	29.4	0.102	20.0	1200	Braid
RG-393	/127-RG393	50.0	ST	29.4	0.390	5.0	5000	Braid
RG-400	/128-RG400	50.0	ST	29.4	0.195	9.6	1900	Braid
RG-401	/129-RG401	50.0	ST	29.4	0.250	4.6	3000	Cu. S-R
RG-402	/130-RG402	50.0	ST	29.4	0.141	7.2	2500	Cu. S-R
RG-403	/131-RG403	50.0	ST	29.4	0.116	29.0	2500	Braid
RG-405	/133-RG405	50.0	ST	29.4	0.086	13.0	1500	Cu. S-R

Note: PE: Polyethylene; ST: Solid Teflon

Using Equation (2.18), the characteristic impedance can be expressed as

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{1}{vC} = \frac{\sqrt{\varepsilon_{re}}}{cC} \tag{2.79}$$

Thus, to compute the characteristic impedance, we just need to obtain the per unit length capacitance C once the effective permittivity is known. This approach makes a difficult task

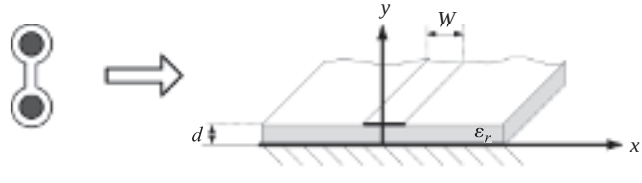


Figure 2.28 Microstrip line

slightly easier. When the thickness of the metal strip can be neglected, it has been found that [2]:

- when $W/d < 1$, the characteristic impedance of the line is

$$Z_0 = \frac{60}{\sqrt{\varepsilon_r}} \ln\left(\frac{8d}{W} + \frac{W}{4d}\right) > \frac{126}{\sqrt{\varepsilon_r}} \tag{2.80}$$

It decreases monotonically to $126/\sqrt{\varepsilon_r}$ as W/d increases to 1.

- when $W/d > 1$, the characteristic impedance of the line is

$$Z_0 = \frac{120\pi}{\sqrt{\varepsilon_r}(W/d + 1.393 + 0.667\ln(W/d + 1.44))} < \frac{126}{\sqrt{\varepsilon_r}} \tag{2.81}$$

It also decreases monotonically from $126/\sqrt{\varepsilon_r}$ as W/d increases. That is, the larger the ratio W/d, the smaller the characteristic impedance; also, the larger the permittivity, the smaller the characteristic impedance. Practical limitations exist on the range of impedances that can be manufactured. These limits depend on factors such as the dielectric constant, substrate height and manufacturing capability. In general, the thinnest line that can be etched routinely with a good photolithographic process is of the order of 0.1 mm. This then puts the upper bound of the impedance at 90–120 Ω. The lower bound is determined by the line width, which should not be comparable to a wavelength. The typical value of the characteristic impedance for industrial standard lines is 50 Ω or 75 Ω.

2.4.3.2 Fundamental Mode

The electromagnetic field distribution around the microstrip line is illustrated by Figure (2.29). Both the electric field and magnetic field are seen to be within the transverse plane. But half of the wave is traveling in free space, which is faster than the other half wave traveling in the substrate, thus this field is in what is called *quasi-TEM mode*, a sort of TEM mode.

A result of a microstrip line being an open structure is that circuits are subject to radiation. This does not mean that they are dangerous to get close to, but the performance of a device or circuit may be affected. This is a direct consequence of the 'unterminated' field lines illustrated in Figure (2.29). In reality, the field lines do not just hang in free space but terminate on whatever is close to the line. The exact relations concerning radiation from a microstrip are complicated but, in general, narrow lines radiate less.

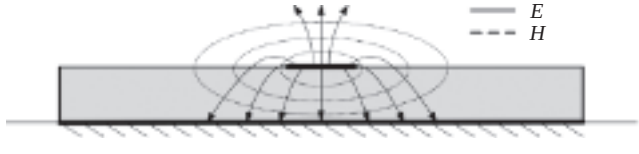


Figure 2.29 The field distribution of a microstrip

The first higher mode in a microstrip line is the transverse electric TE_{10} mode, its cut-off wavelength is twice the strip width. After taking the material and fringing effects into account, the cut-off frequency can be expressed approximately as

$$\lambda_c \approx \sqrt{\varepsilon_r}(2W + 0.8d) \tag{2.82}$$

However, the mode analysis of a microstrip is actually more complicated than this. In addition to the conventional higher order modes, surface modes may exist. The surface mode does not need the metal strip; it only needs the ground plane and the substrate. The lowest transverse electric mode is TE_1 and its cut-off frequency is

$$(f_c)_{TE_1} = \frac{3c\sqrt{2}}{8d\sqrt{\varepsilon_r - 1}} \tag{2.83}$$

The lowest transverse magnetic mode is TM_0 and its cut-off frequency is

$$(f_c)_{TM_0} = \frac{c\sqrt{2}}{4d\sqrt{\varepsilon_r - 1}} \tag{2.84}$$

Obviously $(f_c)_{TE_1} = 1.5 \cdot (f_c)_{TM_0} > (f_c)_{TM_0}$, thus, in order to keep the quasi-TEM mode propagation, the operational frequency of a microstrip line should be smaller than the cut-off frequency of the TE_{10} mode in Equation (2.82) and the cut-off frequency of the TM_0 mode in Equation (2.84). Higher order modes will cause significant power loss via conductive loss and radiation loss. The surface mode may transmit the power to any direction, which is, of course, not desirable.

2.4.3.3 Loss

The loss of a microstrip line comes from the conductor loss and dielectric substrate loss. The radiation loss is negligible at low frequencies. For most microstrip lines, conductor loss is much more significant than dielectric loss. The attenuation constant can be calculated approximately by

$$\alpha_c = \frac{R_s}{Z_0 W} \tag{2.85}$$

where $R_s = \sqrt{\omega\mu/2\sigma}$ is the *surface resistivity* of the conductor.

A summary of some of the common substrates is given in Table 2.4. The first five are hard substrates and the rest are considered soft substrates.

2.4.4 Stripline

A stripline is a conductor sandwiched by a dielectric between a pair of ground planes. It may be viewed as an evolved structure from a coaxial cable, as shown in Figure 2.30. In practice, a stripline is usually made by etching circuitry on a substrate that has a ground plane on

Table 2.4 Some common substrates for a microstrip at 10 GHz

Substrate	ε_r	Loss tangent	Comments
Alumina (Al_2O_3)	9.8	0.0004	Low loss and cost, stable, difficult to machine but very hard-wearing
$LaAlO_3$	24	0.0001	Low loss, but very expensive
MgO	9.8	0.00001	Very low loss but expensive and fragile
Quartz (SiO_2)	3.8	0.0004	Low loss and low permittivity, good for mm range, fragile
Sapphire (Al_2O_3)	9.4 and 10.8	0.00002	Very low loss, single crystal and anisotropic material, expensive
Epoxy (FR4)	4.43@ 1 GHz	0.01	Relatively high loss and low cost, popular PCB, up to $\sim$ 2 GHz
FR2 (flame resistant 2)	4.5@ 1 MHz	0.025	Similar to FR4, cheap, but recommend above $\sim$ 1 GHz
GaAs	13.0	0.0006	Low loss, not cheap, widely used for MMICs
LCP	3.1	0.002	Medium loss, low permittivity, up to 40 GHz, cheap
PTFE (Teflon)	2.1	0.0004	Low loss, medium cost, low permittivity
PTFE-glass	2.1–2.55	$\sim$ 0.001	Medium loss and cost, low permittivity
PTFE-ceramic	10.2	0.002	Medium loss, high permittivity soft substrate, not cheap
RT/Duroid 5870	2.33	0.0012	Medium loss, low cost and low permittivity, up to 40 GHz
RT/Duroid 5880	2.22	0.0009	Low loss, low cost and low permittivity, up to 77 GHz
RT/Duroid 6002, 6202	2.94	0.0012	Medium loss, low cost and low permittivity
RT/Duroid 6006,	6.15	0.0019	Medium loss, low cost and medium permittivity
RT/Duroid 6010	10.2	0.0023	Medium loss, low cost and medium permittivity

the opposite face, then adding a second substrate (which is metalized on only one surface) on top to achieve the second ground plane. The stripline is often considered a 'soft-board' technology, but using low-temperature co-fired ceramics (LTCC), ceramic stripline circuits are also possible.

Unlike a microstrip line, the stripline is basically an enclosed structure; the field is not affected by nearby components. The effective permittivity is the same as the substrate permittivity.

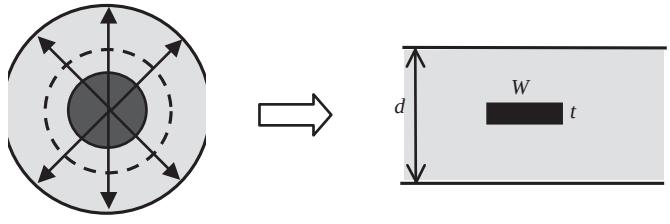


Figure 2.30 From a coaxial cable to a stripline

There are many advantages of using striplines. Whatever circuits are on a microstrip (which is a quasi-TEM mode structure), you can do better by using a stripline, unless you run into fabrication or size constraints. Stripline filters and couplers always offer better bandwidth than their counterparts in a microstrip. Another advantage of the stripline is that fantastic isolation between adjacent traces can be achieved (unlike a microstrip). The best isolation results when a picket-fence of vias surrounds each transmission line, spaced at less than a quarter-wavelength. The stripline can be used to route RF signals across each other quite easily when an offset stripline (i.e. the central conductor is not right at the middle between the two ground planes) is used.

There are two major disadvantages of a stripline:

1. It is much harder and more expensive to fabricate than the microstrip. Lumped-element and active components either have to be buried between the ground planes (not as convenient as a microstrip), or transitions to the microstrip must be employed as necessary to get the components onto the top of the board.
2. Because of the second ground plane, the strip width is much narrower for given impedance (such as 50 ohms) and the board is thicker than that for a microstrip. A common reaction to problems with microstrip circuits is to convert them to a stripline; this may result in a much larger thickness for the same loss of the transmission line.

2.4.4.1 Characteristic Impedance

The characteristic impedance can be calculated approximately by

$$Z_0 = \frac{30\pi}{\sqrt{\varepsilon_r}\left[\dfrac{W}{d-t} + A\right]} \tag{2.86}$$

where

$$A = \left(2B\ln(B+1) - (B-1)\ln(B^2-1)\right)/\pi$$
$$B = 1\Big/\sqrt{1-t/d}$$

The impedance is sensitive to the thickness of the central conductor. The typical value for industrial standard lines is 50 Ω or 75 Ω.

2.4.4.2 Fundamental Mode

Just as for the field within a coaxial cable, the field in a stripline is *TEM mode*. This means that it is nondispersive and the velocity is not changed with frequency.

It is also possible to generate higher order modes if the operational frequency is above the lowest *cut-off frequency*. The smallest wavelength should meet the following condition to avoid higher order modes:

$$\begin{aligned} \lambda_{\min} &> 2d\sqrt{\varepsilon_r} \\ \lambda_{\min} &> 2W\sqrt{\varepsilon_r} \end{aligned} \tag{2.87}$$

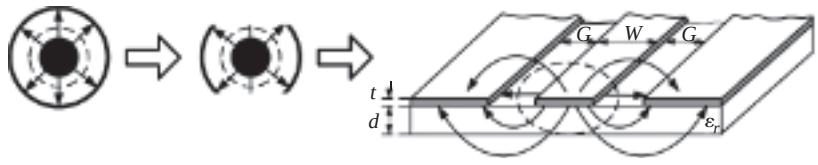


Figure 2.31 Evolution from a coaxial cable to CPW (*G*: gap: *W*: width; *d*: substrate height)

2.4.4.3 Loss

The loss characteristics of the stripline are similar to the microstrip but have little loss due to radiation, as the structure is almost screened.

2.4.5 Coplanar Waveguide (CPW)

The CPW is another popular planar transmission line. Just like a stripline, it may be considered a structure evolved from a coaxial cable, as shown in Figure 2.31. This structure can also be viewed as a coplanar stripline. The central conductor is separated from a pair of ground planes. They all sit on a substrate with a dielectric permittivity of ε. In the ideal case, the thickness of the dielectric is infinite; in practice, it is just thick enough so that EM fields die out before they get out of the substrate. A variant of the coplanar waveguide is formed when a ground plane is provided on the opposite side of the dielectric; this is called a grounded coplanar waveguide (GCPW) and was originally developed to counter the power dissipation problems of CPW.

The CPW offers many advantages, which include the following:

- It is easy to fabricate and to integrate into circuits. Circuit components can be easily mounted on top of the line (even easier than on a microstrip) due to the fact that both the conductor and ground plane are on the same side of the substrate. Unlike the microstrip and stripline, no vias are required.
- It can work to extremely high frequencies (100 GHz or more). Connecting to a CPW does not entail any parasitic discontinuities in the ground plane.
- Good circuit isolation can be achieved using a CPW, because there are always RF grounds between traces. Many examples of high-isolation RF switches have used a grounded CPW to get 60 dB isolation or more.
- The characteristic impedance can be kept as a constant as the signal conductor's width is tapered down/up to meet a pin. This is perfect for matching to a component pin width without changing the substrate thickness.

One disadvantage is potentially lousy heat dissipation - this depends on the thickness of the line and whether it makes contact with a heat sink. In addition, in terms of the circuit size, the CPW is at a disadvantage versus a stripline or microstrip circuit; because its effective dielectric constant is lower (half of the fields are in air). CPW circuits can be lossier than comparable microstrip circuits if a compact layout is required.

2.4.5.1 Characteristic Impedance

The design formulas for a CPW are very complicated. There are four geometric parameters: the gap G, the conductor width W and thickness t and the substrate thickness d. It is not possible to obtain an accurate analytical expression of the characteristic impedance. Some approximations have to be made. If the conductor thickness is neglected, the effective permittivity is given approximately by [2]

$$\varepsilon_{re} = \frac{\varepsilon_r + 1}{2}\left\{ \tanh[0.775\ln(d/G) + 1.75] + \frac{kG}{d}[0.04 - 0.7k + 0.01(1 - 0.1\varepsilon_r)(0.25 + k)]\right\} \tag{2.88}$$

where

$$k = \frac{W}{W + 2G} \tag{2.89}$$

The effective dielectric constant of a CPW is very close to the average dielectric constant of the substrate and free space. One way to think about this is that half of the electric field lines are in free space and half are in the dielectric.

The characteristic impedance is

$$Z_0 = \frac{30\pi}{\sqrt{\varepsilon_{re}}}\frac{K'(k)}{K(k)} \tag{2.90}$$

where $K(k)$ is a complete elliptical function of the first kind. We have

$$k' = \sqrt{1 - k^2};\ K'(k) = K(k') \tag{2.91}$$

and

$$\frac{K'(k)}{K(k)} = \begin{cases} \left[\frac{1}{\pi}\ln(2\frac{1+\sqrt{k'}}{1-\sqrt{k'}})\right] & if \quad 0 < k < 0.707 \\ \left[\frac{1}{\pi}\ln(2\frac{1+\sqrt{k}}{1-\sqrt{k}})\right]^{-1} & if \quad 0.707 < k < 1 \end{cases} \tag{2.92}$$

Again, the typical impedance value for industrial standard lines is 50 Ω or 75 Ω.

2.4.5.2 Fundamental Mode

The EM field distribution around the CPW is illustrated in Figure 2.31 and is similar to that of a microstrip. The wave velocity in the air is faster than that in the substrate, thus the fundamental field of CPW is *quasi-TEM mode*.

Higher order modes and surface modes may be generated in a CPW just as in a microstrip line. Thus, ground straps (bounding wires) are normally needed to tie the two grounds together in a CPW. These are especially important around any discontinuity, such as a tee junction. Care has to be taken since the bounding wires themselves could be the cause of discontinuity!

2.4.5.3 Loss

The current on the CPW is concentrated around the signal conductor. The current on the ground planes is also very focused in a small area, which results in a relatively high conductor loss as well as a heat dissipation problem. Generally speaking, the CPW exhibits a higher loss than its microstrip counterpart.

2.4.6 Waveguide

This is a very special and unique electromagnetic transmission line. Unlike any other transmission lines, a waveguide consists of just one piece of metal, which is tubular, usually with a circular or rectangular cross-section. A rectangular waveguide is shown in Figure 2.32. Due to the boundary conditions that the electric and magnetic fields have to satisfy, there are many possible wave patterns, which are called *transverse electric* (m, n) *modes* (TE_{mn} modes) and *transverse magnetic* (m, n) *modes* (TM_{mn} modes). m and n represent the number of peaks along the x and y axes respectively. For example, TE_{10} mode means that the electric field is within the transverse plane and there is no electric field component along the propagation direction, whilst the magnetic field is not confined to the transverse plane, and the electric field changes along the x-axis, having one peak, but it has no changes along the y-axis. Which modes will actually be generated inside a waveguide depend on the frequency and excitation.

If the frequency is below the cut-off frequency of the fundamental mode (which is TE_{10} mode for standard waveguides), no propagation mode can be generated. Thus, the operational frequency should be greater than the cut-off frequency, which means that the waveguide can be

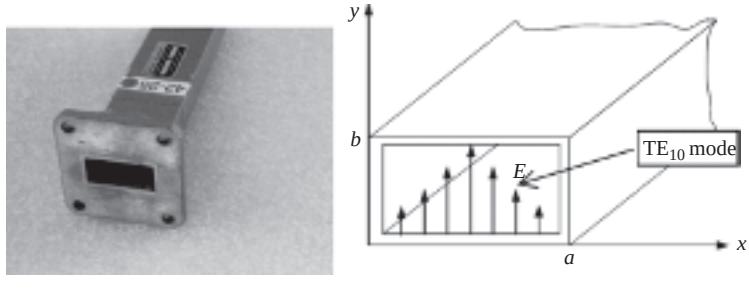


Figure 2.32 Rectangular waveguide

considered a high-pass filter. Any signals below the cut-off frequency will be filtered out by the waveguide. The larger the waveguide, the lower the cut-off frequency. Since large waveguides are heavy and expensive, they are not attractive for applications. Thus, waveguides are only used for microwave and millimeter wave frequency bands.

The main advantages of the waveguide are low loss and high power-handling capacities, which are very important for high-power applications such as radar.

2.4.6.1 Fundamental Mode

The fundamental mode of a standard rectangular waveguide is TE_{10} mode. The field pattern (along with some higher modes) is illustrated in Figure 2.32 (the field patterns for some higher modes can be found in [2] and [4]). The width and height of the waveguide are a and b respectively. The electric field can be expressed as

$$\begin{aligned} E_y &= E_0 \sin(\tfrac{\pi}{a}x)e^{j(\omega t-\beta z)} \\ E_x &= E_z = 0 \end{aligned} \tag{2.93}$$

and the magnetic field is given by

$$\begin{aligned} H_x &= H_1 \sin(\tfrac{\pi}{a}x)e^{j(\omega t-\beta z)} \\ H_y &= 0 \\ H_z &= H_2 \cos(\tfrac{\pi}{a}x)e^{j(\omega t-\beta z)} \end{aligned} \tag{2.94}$$

The electric field is indeed within the transverse plane and the maximum field is at $x = a/2$, while the magnetic field has two components, one along the propagation direction z. Neither the electric nor the magnetic field is a function of y, i.e. the mode index $n = 0$.

2.4.6.2 Cut-off Frequency, Waveguide Wavelength and Characteristic Impedance

The cut-off wavelength for TE_{mn} and TM_{mn} modes is given by

$$\lambda_c = \frac{2}{\sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2}} \tag{2.95}$$

and its corresponding cut-off frequency is

$$f_c = \frac{1}{2\sqrt{\varepsilon\mu}}\sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2} \tag{2.96}$$

Thus, for TE_{10} mode, the cut-off wavelength is $\lambda_c = 2a$. This means that the waveguide works for the operational wavelength $\lambda < 2a$.

Since the next highest mode to TE_{10} is TE_{20} (for a standard waveguide, $2b < a$, i.e. TE_{01} is a higher mode than TE_{20}) and its cut-off wavelength is $\lambda_c = a$, this waveguide is only suitable for the operational wavelength between these two cut-off wavelengths, i.e. $a < \lambda < 2a$. Outside

Table 2.5 Standard waveguides

Waveguide	Freq. (GHz)	ID of *a* (mm)	ID of *b* (mm)	Freq. band
WR-137	5.85–8.2	34.85	15.80	C band
WR-112	7.05–10.00	28.50	12.60	H band
WR-90	8.2–12.4	22.86	10.16	X band
WR-62	12.4–18.0	15.80	7.90	Ku band
WR-51	15.0–22.0	12.96	6.48	K band
WR-42	18.0–26.5	10.67	4.32	K band

the bounds, the frequency is either too low for transmission or too high to keep single-mode transmission. Higher modes are not desirable since they have higher loss and the field pattern may be changed over the transmission. Thus, each industrial standard waveguide is only suitable for a certain frequency range. A list of some selected standard waveguides with their suitable frequency ranges and inside dimensions is given in Table 2.5.

The field inside a waveguide exhibits a periodic feature and the period is one *waveguide wavelength*, which is actually different from (longer than) the free space wavelength. The waveguide wavelength can be calculated using

$$\lambda_g = \lambda \Big/ \sqrt{1 - (\frac{\lambda}{\lambda_c})^2} \tag{2.97}$$

It is determined by the free space wavelength λ and the cut-off frequency λ_c.

The characteristic impedance is also mode-dependent. For TE_{10} mode, it is

$$Z_{TE10} = 120\pi \Big/ \sqrt{1 - (\frac{\lambda}{2a})} \tag{2.98}$$

which is not a constant but a function of frequency. This is one of the reasons why the single mode is preferred for waveguide applications.

In this section six of the most popular transmission lines have been introduced and discussed. The characteristic impedance, fundamental mode and loss characteristics have been presented. A brief summary is provided in Table 2.6.

2.5 Connectors

In practice, almost all transmission lines have to be terminated to suitable connectors, which make the device interconnection much easier. There are many types of industrial standard connectors. Sometimes, more than one connector is available. For example, RG58 cable can be assembled to SMA, SMB, BNC and Type N connectors. Making the right choice can be a problem and there is a lack of information on this practical subject in other books, hence we will address this issue in this section.

Connectors are developed as a pair: a male and a female (an example is shown in Figure 2.33), although some RF connectors are sexless (such as the APC-7 and the

Table 2.6 Summary of various transmission lines

	Two-wire	Coax	Microstrip	Stripline	CPW	Wave-guide
Basic mode	TEM	TEM	quasi-TEM	TEM	quasi-TEM	TE_{10}
Bandwidth	narrow – low freq	wide	wide	wide	very wide	narrow
Loss	high	med.	med.	med.	med.	low
Cost	low	low/ med.	med.	med.	med.	high
Ease of integration	med.	hard	easy	med.	easy	hard
Application	low freq., short distance	general purpose	PCB circuit and MMIC	RF circuit and MMIC	RF circuit and MMIC	high power and high freq.

General Radio GR874). It is very important to choose the right connector for the application, since the effects of the connector (which is an additional element and may not have been taken into account in the design) on the system performance and measurements may be quite significant. This is especially true in antenna measurements.

A good summary of all RF/microwave connectors was given by Amphenol RF [5]. Table 2.7 is a selection of some popular connectors for antenna systems and measurements. These connectors are mainly for coaxial cables, but some of them can be used for other transmission lines. For example, the SMA straight PCB mount jack, tab terminal panel jack and stub terminal panel jack are available and widely used in the antenna community.

Since all RF test equipment comes with coaxial connectors (type N and SMA are popular connectors), direct connection with other forms of transmission lines (such as microstrip and CPW) would be tricky. Some adapters have been developed. For example, industrial standard coax-to-waveguide adaptors are now widely available on the market. Figure 2.34 shows how to connect an SMA connector directly to a microstrip and CPW (feed line to an antenna) in practice – no standard adaptor is available.

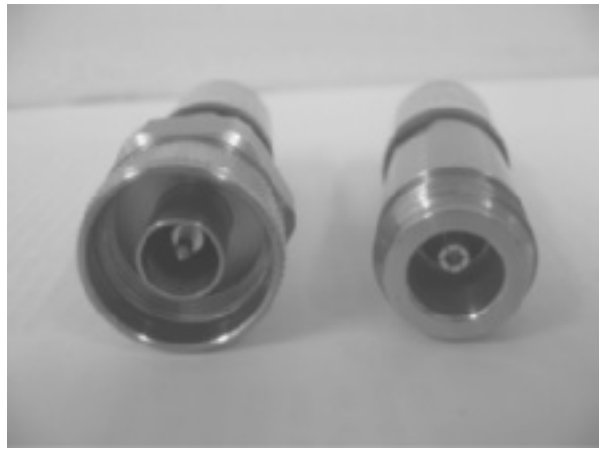


Figure 2.33 Male (left) and female (right) N-type connectors

Table 2.7 Some industry standard connectors (Reproduced by permission of Amphenol RF)

	BNC Baby N connector (BNC). Bayonet-style coupling for quick connection and disconnection. Available in 50 Ω, 75 Ω and 50 Ω reverse polarity. DC – 4 GHz
	K The K Connector™ is a precision coaxial connector system that operates up to 40 GHz. It is compatible with SMA, WSMA and 3.5 mm connectors. It is well suited to applications in components, systems or instrumentation.
	MCX A snap-on miniature coaxial (MCX) connector that conforms to the European CECC 22220. Since the MCX has identical inner contact and insulator dimensions to the SMB while being 30% smaller, it provides designers with options where weight and physical space are limited. DC – 6 GHz.
	MMCX A micro-miniature coaxial (MMCX) connector with a lock-snap mechanism allowing for 360 degrees rotation on a printed circuit board. Conforms to the European CECC 22000 specification and comes in surface mount, edge card and cable connectors. DC – 6 GHz.
	Precision: APC-2.4, APC-3.5, APC-7 and APC-N The acronym APC describes high-performance precision connectors, and stands for Amphenol Precision Connectors. Developed by Amphenol and Hewlett-Packard engineers. DC – 50 GHz.
	Mini BNC A new generation of miniature BNC connectors that maintain the positive characteristics of our full-size BNCs for 75 Ω systems while allowing 40% more interconnects in the same area. DC – 11 GHz.
	SMA Subminiature version A (SMA) connectors with a threaded coupling mechanism that perform through 18 GHz. Available in standard, phase adjustable and reverse polarity. Built in accordance with MIL-C-39012 and CECC 22110/111, SMA connectors can be mated with all connectors that meet these specs, regardless of manufacturer. Widely used with RG-55, 58, 141, 142, 223, 303, 122, 174, 188 and 316.
	SMB Subminiature version B (SMB) connectors. Developed in the 1960s as a smaller alternative to the SMA, the SMB line features a snap-on coupling mechanism. Available in 50 Ω, 75 Ω and miniature 75 Ω. DC – 4 GHz (usable to 10 GHz). Often used with RG-188 and 196.

(continued)

Table 2.7 Some industry standard connectors (Reproduced by permission of Amphenol RF) (*Continued*)

	SMC Subminiature version C (SMC) connectors. Medium-sized 50 Ω threaded connectors designed to meet MIL-C-39012 category D, as generated by the US Air Force. DC – 4 GHz (usable to 10 GHz). Often used with RG-188 and 196.
	SMP Subminiature connectors with a frequency range up to 40 GHz. Used in miniaturized applications they feature both push-on and snap-on mating styles.
	SSMB Scaled SMA (SSMA). Microminiature connectors with snap-on mating interface allowing quick installation in small spaces with excellent performance in devices up to 4 GHz.
	TNC Features screw threads for mating and serves as a threaded version of the BNC connector. The TNC is a 50 Ω connector available in both standard and reverse polarity. DC – 11 GHz.
	Type N Available in standard N (coaxial cable) and corrugated N (helical and annular cable), the Type N is a durable, weatherproof, medium-sized connector consistent through 11 GHz. Used with RG-8, 58, 141 and 225.
	UHF Invented for use in the radio industry, UHF stands for ultra-high frequency. While at the time 300 MHz was considered high frequency, these are now general purpose connectors for low-frequency systems.

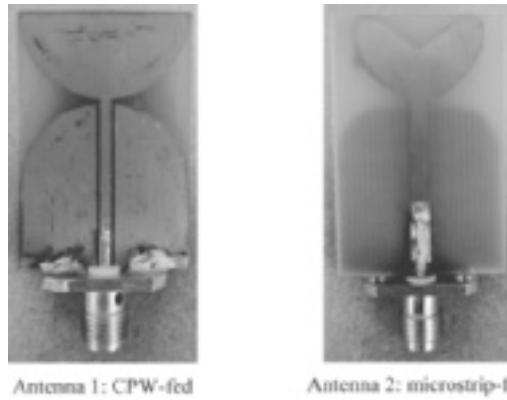


Figure 2.34 Wideband antennas fed by CPW and microstrip, which are directly connected/soldered to SMA connectors

2.6 Summary

This chapter has provided comprehensive coverage of circuit concepts and transmission lines – this is essential knowledge for antenna feeding, matching and characterization. In summary:

- An introduction to lumped element systems and distributed element systems has been given right at the beginning. The main idea is that the current, voltage and impedance are all functions of the frequency and the reference position at the transmission line.
- A transmission line model has been developed to obtain the important parameters of a transmission line, which include the characteristic impedance, input impedance, attenuation constant, phase constant and velocity. An extensive study on terminated transmission lines has been carried out. The reflection coefficient, return loss and VSWR have been introduced to evaluate the line impedance matching.
- The Smith Chart has been introduced as a very useful tool to analyze impedance matching. Lumped and distributed matching networks and impedance-matching techniques have also been addressed.
- The bandwidth and quality factor (Q factor) have been discussed in depth.
- Six popular transmission lines have been examined and compared in terms of their characteristic impedance, fundamental mode, loss characteristics and frequency bandwidth.
- Various RF/microwave cables and connectors have been presented at the end of this chapter along with their typical specifications and frequency bandwidths.

References

[1] J. D. Kraus and D. A. Fleisch, *Electromagnetics with Applications*, 5th edition, McGraw-Hill, Inc., 1999.
[2] D. M. Pozar, *Microwave Engineering*, 2nd edition, John Wiley & Sons, Inc., 1997.
[3] http://www.rfcafe.com/references/electrical/coax_chart.htm
[4] C. S. Lee, S. W. Lee and L. L. Chuang, 'Plot of modal field distribution in rectangular and circular waveguides', *IEEE Trans. on MTT*, March, pp. 271–274, 1985.
[5] www.amphenolrf.com

Problems

Q2.1 Explain the concept of the characteristic impedance of a transmission line.

Q2.2 For a low-loss transmission line, find its characteristic impedance, attenuation constant and phase constant (or wave number using lumped elements). How does the frequency affect these parameters?

Q2.3 A uniform transmission line has constants $R = 500\mu\Omega/m$, $G = 1.5mS/m$, $L = 0.5\mu H/m$ and $C = 10nF/m$. Find the characteristic impedance and the attenuation constant of the line at the following frequencies

a) 50 Hz;
b) 30 MHz;
c) 1 GHz;
d) 10 GHz.

and comment on the results.

Q2.4 A coaxial transmission line has $a = 4$ mm and $b = 12$ mm. Find the characteristic impedance of the line if the dielectric is
a) air space polyethylene (ASP);
b) foam polyethylene (FE);
c) solid Teflon (ST).
Hint: Use Table 2.2 for permittivity.

Q2.5 A 100 Ω resistor is connected to a good cable with characteristic impedance of 50 Ω. The attenuation constant is not zero but 0.2 Np/m at 1 GHz, and the relative permittivity of the cable dielectric is 1.5. If the cable length is 10 m, find
a) the reflection coefficient and return loss at the termination;
b) the reflection coefficient and return loss at the input of the cable;
c) the VSWR at both the terminal and the input of the cable;
d) the input impedance at the input of the cable.
Suggest a method to improve the matching of the system.

Q2.6 Obtain the theoretical characteristic impedance and the best impedance for maximum power-handling capacity of a coaxial cable. Use the results to justify why the most common coaxial cable impedances are 50 Ω and 75 Ω.

Q2.7 RG-59U, a popular cable for microwave applications, has an open-circuit impedance of $130 + j75$ ohms and short-circuit impedance of $30.3 - j21.2$ ohms. Find the characteristic impedance of the line.

Q2.8 A quality transmission line is terminated in $100 + j50$ ohms. Find
a) the voltage reflection coefficient;
b) the VSWR;
c) the shortest length of line required to transform the impedance to purely resistive.
If 220 V is applied to the line, find the maximum and minimum line voltages.

Q2.9 Explain the concept of impedance matching and then compare a lumped matching network and a distributed matching network.

Q2.10 Explain what the Smith Chart is. Explain its application.

Q2.11 A load with an impedance of $100 - j100$ Ω is to be matched with a 50 Ω transmission line. Design a matching network and discuss if there are other solutions available.

Q2.12 A load with an impedance of $100 - j100$ Ω is to be matched with a 50 Ω transmission line. Design two stub-matching networks and then compare their bandwidth performance.

Q2.13 Explain what Bode–Fano limits are and how they may be applied to matching networks.

Q2.14 Discuss the relationship between bandwidth and quality factor. What is the major difference between the loaded Q factor and the unloaded Q factor?

Q2.15 Design a 50-ohm microstrip line using a PCB board with PTFE (Teflon) substrate of 1 mm thickness. Find the cut-off frequency for the first higher mode in the line.

Q2.16 Design a 50-ohm CPW using a PCB board with PTFE (Teflon) substrate of 1 mm thickness.

Q2.17 Rectangular waveguides are widely used for radar applications. WR-90 standard waveguide (see Table 2.5) is mainly used for the X band. Find
a) the cut-off frequency for TE_{10} mode;
b) the cut-off frequency for TE_{01} mode;
c) the cut-off frequency for TE_{20} mode.
Hence identify the most suitable frequency range for this waveguide.

3

Field Concepts and Radio Waves

In this chapter we will first see how Maxwell's equations can be used to obtain wave solutions. The concepts of the plane wave, intrinsic impedance and polarization will then be introduced, followed by a discussion on radio propagation mechanisms and radio wave propagation characteristics in various media. A few basic radio propagation models will be introduced, and circuit concepts and field concepts will be compared at the end of this chapter. The concept of skin depth will be looked into from both the field and circuit points of view. Although the issues addressed in this chapter may not be used directly for antenna design, the knowledge will be extremely useful for gaining a better understanding of the antenna radiation characteristics as well as radio waves – generated/received by antennas. Because antennas and radio propagation are so closely linked, some countries and universities treat them as a single subject.

3.1 Wave Equation and Solutions

As mentioned in Chapter 1, Maxwell's modified version of Ampere's Circuital Law enables a set of equations to be combined together to derive the electromagnetic wave equation. The derivation is relatively straightforward.

Now let us discuss a time-harmonic case with the time factor $e^{j\omega t}$, which means a single frequency and is the most common form of a wave in real life (according to Fourier's theory, more complicated cases may be decomposed to a linear combination of harmonic waves). From Maxwell's equations (1.29), we have

$$\begin{aligned} \nabla \times \boldsymbol{E} &= -j\omega\mu \boldsymbol{H} \\ \nabla \times \boldsymbol{H} &= (\sigma + j\omega\varepsilon)\boldsymbol{E} \\ \nabla \bullet \boldsymbol{E} &= \rho/\varepsilon \\ \nabla \bullet \boldsymbol{H} &= 0 \end{aligned} \tag{3.1}$$

Take a curl operation on the first equation to yield

$$\nabla \times \nabla \times \boldsymbol{E} = \nabla(\nabla \bullet \boldsymbol{E}) - \nabla^2 \boldsymbol{E} = -j\omega\mu \nabla \times \boldsymbol{H}$$

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

where $\nabla^2 = \nabla \bullet \nabla = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2} + \dfrac{\partial^2}{\partial z^2}$.
Combine this with the second and third equations in Equation (3.1) to obtain

$$\nabla^2 \mathbf{E} - j\omega\mu(\sigma + j\omega\varepsilon)\mathbf{E} = \nabla(\rho/\varepsilon) \tag{3.2}$$

Now let

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\varepsilon)} = \alpha + j\beta \tag{3.3}$$

where α and β are the *attenuation constant* and *phase constant*, respectively. Similar definitions were introduced for a transmission line in Chapter 2. From Equation (3.3) we can represent these constants by the material properties and frequency as:

$$\begin{aligned} \alpha &= \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\left(\sqrt{1 + \frac{\sigma^2}{\varepsilon^2\omega^2}} - 1\right)\right]^{1/2} \\ \beta &= \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\left(\sqrt{1 + \frac{\sigma^2}{\varepsilon^2\omega^2}} + 1\right)\right]^{1/2} \end{aligned} \tag{3.4}$$

Equation (3.2) can now be rewritten as

$$\nabla^2 \boldsymbol{E} - \gamma^2 \boldsymbol{E} = \nabla(\rho/\varepsilon) \tag{3.5}$$

In the source-free region ($\rho = 0$), we have

$$\nabla^2 \boldsymbol{E} - \gamma^2 \boldsymbol{E} = 0 \tag{3.6}$$

This is called the *wave equation*. There are many possible solutions to this equation. Boundary conditions and sources are required to obtain the specific solutions. In free space, one of the solutions is

$$\boldsymbol{E} = \hat{\boldsymbol{x}} \boldsymbol{E}_0 e^{j\omega t - \gamma z} = \hat{\boldsymbol{x}} \boldsymbol{E}_0 e^{-\alpha z + j(\omega t - \beta z)} \tag{3.7}$$

This can be validated easily by using this representation in the wave equation. Other possible solutions include, for example

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_0 e^{j\omega t + \gamma z};\ \boldsymbol{E} = \hat{\boldsymbol{y}} E_0 e^{j\omega t + \gamma z};\ \boldsymbol{E} = \hat{\boldsymbol{z}} E_0 e^{j\omega t \pm \gamma x};\ \ldots$$

Using the electric field $\boldsymbol{E}$ in Equation (3.7) and Equation (3.1), the magnetic field $\boldsymbol{H}$ is

$$\boldsymbol{H} = \frac{j}{\omega\mu}\nabla \times \boldsymbol{E} = -\hat{\boldsymbol{y}}\frac{j\gamma}{\omega\mu} E_0 e^{-\alpha z + j(\omega t - \beta z)} \tag{3.8}$$

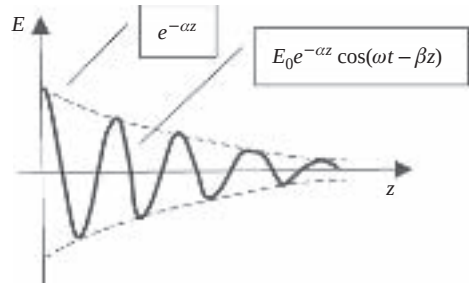


Figure 3.1 A traveling wave in a medium with loss

Thus, the magnetic field has only a *y* component in this case, which is orthogonal to the electric field. There is a phase difference between the electric and magnetic fields if the attenuation constant is not zero.

3.1.1 Discussion on Wave Solutions

Equation (3.7) can be illustrated by Figure 3.1; it is evident that

- The wave solution is a vector. In this case, it has only an x component.
- Its amplitude is decreased exponentially as a function of the propagation distance (z, in this case). The attenuation constant α, given by Equation (3.4), is determined by the material properties and frequency. When the conductivity σ is zero, the wave amplitude is a constant.
- Its phase φ is of the form $(\omega t - \beta z)$, which is a function of time, frequency and propagation distance.

For the loss-free case, $\sigma = 0$, Equation (3.4) can be simplified to

$$\begin{aligned} \alpha &= 0 \\ \beta &= \omega\sqrt{\mu\varepsilon} \end{aligned} \tag{3.9}$$

If we fix the phase and let the wave travel a distance of Δz over a period of time Δt, mathematically this is

$$\Delta\varphi = \omega \cdot \Delta t + \beta \cdot \Delta z = 0 \tag{3.10}$$

Thus, the velocity of the wave can be obtained as

$$v = \frac{\Delta z}{\Delta t} = \frac{\omega}{\beta} \tag{3.11}$$

Replace β by Equation (3.9) to give

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \tag{3.12}$$

This means that the wave velocity is determined by the permittivity and permeability of the medium in which the wave is traveling. In free space this velocity is

$$v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \approx 3 \times 10^8 \text{ m/s}$$

We can therefore conclude that the velocity of an electromagnetic wave (including light) in free space is about 3×10^8 m/s – this was what Maxwell obtained more than 120 years ago when he formulated the four equations (more precisely 20 equations in his original work), but at that time nobody could validate this important result. This is a good example of how mathematics can be used to solve real world engineering problems.

In addition, from Equation (3.11), we can see that

$$\beta = \frac{\omega}{v} = \frac{2\pi f}{v} = \frac{2\pi}{\lambda} \tag{3.13}$$

Thus, the phase constant is also called the *wave number* (for every one wavelength, the phase is changed by 2π), which is the same as what we obtained for a transmission line in Chapter 2. In fact, the transmission line equation (2.7) is just a special case of the wave equation (3.6) when x and y are fixed. The free space could be viewed as an open transmission line where the information is carried by electromagnetic (EM) waves.

Light, X-rays and radio waves are EM waves at different frequencies although they seem to be very different. One thing that all the forms of EM waves have in common is that they can travel through empty space. This is not true for other kinds of waves; sound waves, for example, need some kind of material, like air or water, in which to move. EM energy is carried by photons, the energy of a photon is hf, where h is Planck's constant $= 6.63 \times 10^{-34}$ Js, and f is the frequency in Hz.

3.2 The Plane Wave, Intrinsic Impedance and Polarization

3.2.1 The Plane Wave and Intrinsic Impedance

When the conductivity of the medium is zero, the electric field in Equation (3.7) can be simplified to

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_0 e^{j(\omega t - \beta z)} \tag{3.14}$$

and the corresponding magnetic field is

$$\boldsymbol{H} = \hat{\boldsymbol{y}} \frac{\beta}{\omega \mu} E_0 e^{j(\omega t - \beta z)} = \hat{\boldsymbol{y}} \sqrt{\frac{\varepsilon}{\mu}} E_0 e^{j(\omega t - \beta z)} \tag{3.15}$$

Thus, the electric and magnetic fields are in phase, of constant amplitude and orthogonal to each other, as well as being orthogonal to the propagation direction z. This EM wave is called the *plane wave*, and is illustrated by Figure 3.2. It is a special but common form of EM wave whose amplitude is a constant (in theory). The spherical wave is another common wave form.

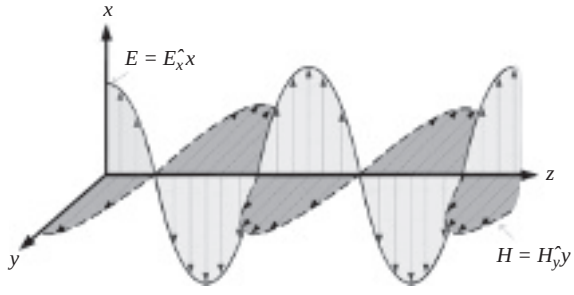


Figure 3.2 A plane wave traveling in the z-direction

The plane wave can only exist far away from the source, whilst the spherical wave is considered closer to the source.

The *power flow density* of the EM wave, also known as the *Poynting vector*, is defined as the cross product of the electric and magnetic fields, i.e.

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}^* \,(\mathrm{W/m^2}) \tag{3.16}$$

where * denotes the complex conjugate, i.e. $(R + jX)^* = R - jX$. The Poynting vector describes the amplitude and direction of the flow of power density in EM waves. It is named after the English physicist John Henry Poynting, who introduced it in 1884. The power flow direction is orthogonal to $\boldsymbol{E}$ and $\boldsymbol{H}$. Equation (3.16) gives the instantaneous Poynting vector. The averaged Poynting vector is obtained by integrating the instantaneous Poynting vector over one period and dividing by one period. Thus, the time-averaged power density of an EM wave is

$$\boldsymbol{S}_{av} = \frac{1}{2}\mathrm{Re}(\boldsymbol{E} \times \boldsymbol{H}^*) = \hat{\boldsymbol{z}}\frac{1}{2}\sqrt{\frac{\varepsilon}{\mu}}E_0^2 \tag{3.17}$$

and it can be obtained by its electric field amplitude and material properties $\sqrt{\varepsilon/\mu}$ – this is actually the ratio of the electric field to the magnetic field:

$$\eta = \frac{E}{H} = \sqrt{\frac{\mu}{\varepsilon}} = 120\pi\sqrt{\frac{\mu_r}{\varepsilon_r}}\,(\Omega) \tag{3.18}$$

It is called the *intrinsic impedance* of the material and is determined by the ratio of the permittivity to the permeability of the medium for a loss-free medium. If the medium is lossy (conductivity is not negligible and/or the permittivity is complex), the intrinsic impedance is complex:

$$\eta = \frac{E}{H} = -\frac{\omega\mu}{j\gamma} = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\varepsilon}}\,(\Omega) \tag{3.19}$$

In free space it is

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi \approx 377\,(\Omega) \tag{3.20}$$

and the time-averaged power density is

$$\boldsymbol{S}_{av} = \hat{\boldsymbol{z}}\frac{1}{2}\eta_0 E_0^2 = \hat{\boldsymbol{z}}60\pi E_0^2 \tag{3.21}$$

3.2.2 Polarization

A very important feature of the EM wave is the *polarization*, which is described by the locus of the tip of the $\boldsymbol{E}$ vector as time progresses. If we use a trigonometric form (we can also use the exponential form), a wave propagating towards the z direction can be expressed as

$$\boldsymbol{E} = \hat{\boldsymbol{x}}A\cos(\omega t - \beta z) + \hat{\boldsymbol{y}}B\sin(\omega t - \beta z) \tag{3.22}$$

where A and B, also shown in Figure 3.3, are the amplitudes of the field components in the x and y directions, respectively. It is not difficult to verify that this $\boldsymbol{E}$ field is also a solution of the wave equation (3.6).

If A or $B = 0$, this expression represents a *linearly polarized* wave; if $A \neq B \neq 0$, it is an *elliptically polarized* wave; if $A = B$, it then represents a *circularly polarized* wave, which is widely employed in satellite communications. Because the ionosphere causes *Faraday rotation* to an EM wave, which means that a linearly polarized EM wave may be rotated by an unknown amount (depending on the thickness and temperature of the ionosphere, as well as the frequency – the rotation is high at lower frequencies but small at higher frequencies), making the linearly polarized wave hard to match after passing through the ionosphere. However, there is no problem for circularly polarized waves; this is why satellite systems like GPS (global positioning system) have employed circular polarization, not linear polarization, for transmission.

The circularly polarized wave may be considered a combination of two linearly polarized waves. There are two types of circular polarization: *right-hand circular polarization* (RCP)

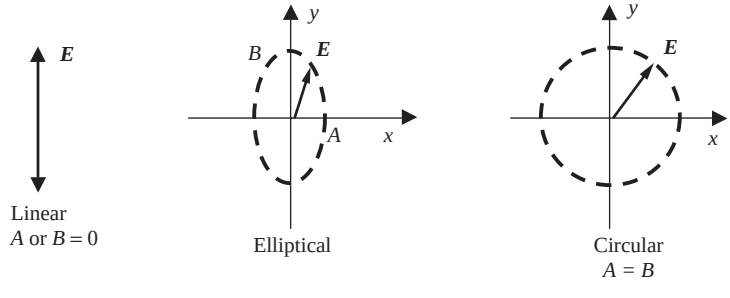


Figure 3.3 Wave polarizations

and *left-hand circular polarization* (LCP) – one linearly polarized wave is ahead of or behind the other one by 90 degrees. When the thumb points to the propagation direction, if the tip of the $\boldsymbol{E}$ vector follows the right-hand fingers as time progresses, it is RCP. Otherwise, it is LCP. Equation (3.22) represents a right-hand circularly polarized wave if $A = B > 0$. Its corresponding left-hand polarized wave can be expressed as

$$\boldsymbol{E} = \hat{\boldsymbol{x}} A \cos(\omega t - \beta z) - \hat{\boldsymbol{y}} B \sin(\omega t - \beta z) \tag{3.23}$$

There is just a sign change: '+' is changed to '−' for the y component. The ratio of amplitudes A to B is called the *axial ratio*:

$$AR = \frac{A}{B} \tag{3.24}$$

For a circularly polarized wave, *AR* is one. For a linearly polarized wave, it is infinite or zero, thus $0 \leq AR \leq +\infty$.

It should be pointed out that a plane wave can be linearly polarized, circularly polarized or elliptically polarized. Equation (3.14) represents a linearly polarized plane wave, whilst Equations (3.22) and (3.23) are circularly polarized plane waves.

3.3 Radio Wave Propagation Mechanisms

Radio wave propagation is a special subject. A radio wave is considered a general term in this book for EM waves up to about 100 GHz. In this section we are going to briefly review wave propagation mechanisms, which include wave reflection, transmission, diffraction and scattering.

3.3.1 Reflection and Transmission

As we understand now, an EM wave far away from its source may be considered a local plane wave. Let a linearly polarized plane wave be incident on the surface between Medium 1 and Medium 2, as shown in Figure 3.4. What is going to happen at the boundary? The wave will be partially reflected back to Medium 1 and partially transmitted (more precisely refracted; the *refraction* is the change in direction of a wave due to a change in velocity from one medium to another) into Medium 2. If the loss can be neglected, there are a few important points to note:

- the incident angle θ_i is the same as the reflected angle θ_r, that is $\theta_i = \theta_r$;
- the incident angle θ_i is linked to the transmitted angle θ_t by *Snell's law*:

$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{\gamma_1}{\gamma_2} = \frac{\sqrt{\varepsilon_1\mu_1}}{\sqrt{\varepsilon_2\mu_2}} \tag{3.25}$$

The *reflection coefficient* is defined as the ratio of the reflected wave to the incident wave, i.e.

$$\Gamma = \frac{E_r}{E_i} \tag{3.26}$$

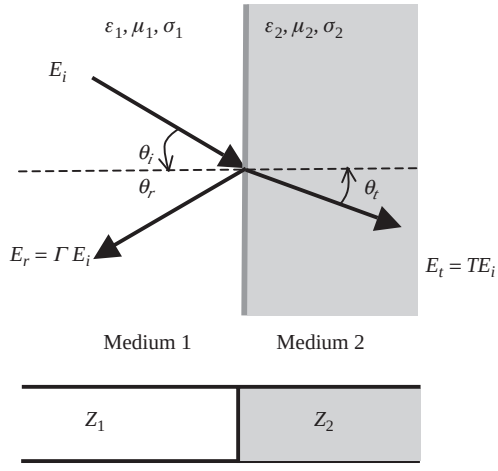


Figure 3.4 Plane wave reflection and transmission, and its analogous transmission line

and the *transmission coefficient* is defined as the ratio of the transmitted wave to the incident wave, i.e.

$$T = \frac{E_t}{E_i} \tag{3.27}$$

Both coefficients are linked to the wave polarization. There are basically two orthogonal polarizations – parallel polarization (***E*** is parallel to the incident plane formed by the incident and reflected waves) and perpendicular polarization (***E*** is perpendicular to the incident plane), as shown in Figure 3.5. Any other polarizations can be considered to be combinations of these two principal polarizations.

We can employ either the field approach or the circuit approach to obtain the reflection and transmission coefficients. Using field concepts, we need to employ the boundary conditions, which is relatively complicated. Thus, we are going to use the circuit approach to obtain these coefficients.

From Figure 3.4, we can see that the two media can be replaced by two analogous transmission lines with characteristic impedances of Z_1 and Z_2, respectively. They are determined by

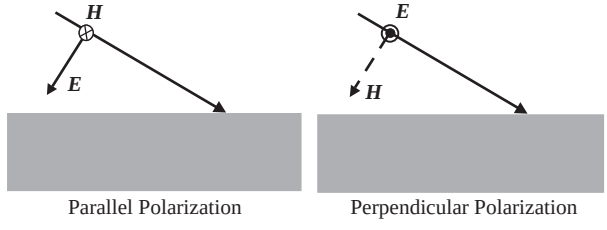


Figure 3.5 Two principal polarizations

wave polarization, incident angle and material properties:

$$Z_1 = \begin{cases} \eta_1 \cdot \cos\theta_i; & \text{for parallel pol.} \\ \eta_1 / \cos\theta_i; & \text{for perpendicular pol.} \end{cases}$$
$$Z_2 = \begin{cases} \eta_2 \cdot \cos\theta_t; & \text{for parallel pol.} \\ \eta_2 / \cos\theta_t; & \text{for perpendicular pol.} \end{cases} \tag{3.28}$$

where the intrinsic impedance η is determined by material properties, as shown in Equation (3.19). The reflection coefficient between these two transmission lines can be obtained easily from Equation (2.28), i.e.:

$$\Gamma = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{3.29}$$

which is also the reflection coefficient at the boundary between Medium 1 and Medium 2. The transmission coefficient is

$$T = 1 + \Gamma = \frac{2Z_2}{Z_2 + Z_1} \tag{3.30}$$

The reflection and transmission coefficients are ratios of electric field strengths. Because the incident power equals the sum of the reflected and transmitted powers, we have:

$$|\Gamma|^2 + |T|^2 = 1 \tag{3.31}$$

Example 3.1: Reflection on a perfect conductor. Obtain the reflection and transmission coefficients between air and a perfect conductor.

Solution:
The conductivity of a perfect conductor is infinite. Using Equations (3.19) and (3.28), we know that the characteristic impedance of its equivalent transmission line is zero for any polarization and incident angle, i.e. $Z_2 = 0$, thus

$$\Gamma = -1 \text{ and } T = 0$$

This means that all signals are reflected back and the phase is changed by 180 degrees. There is no signal transmitted into the conductor.

Example 3.2: Reflection on a ground. If the relative permittivity of a ground is 9 and the conductivity is very small and negligible, plot the reflection coefficient as a function of the incident angle for both parallel and perpendicular polarizations.

Solution:

$\varepsilon_1 = 1$ and $\varepsilon_2 = 9$. To obtain the reflection coefficient, we first use Equation (3.25) to yield

$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{\sqrt{\varepsilon_1\mu_1}}{\sqrt{\varepsilon_2\mu_2}} = \frac{1}{3}, \text{ thus } \cos\theta_t = \sqrt{1-\sin^2\theta_t} = \sqrt{1-\frac{1}{9}\sin^2\theta_i};$$

and then use Equation (3.28) to give

$$Z_1 = \begin{cases} 120\pi \cdot \cos\theta_i; \text{ for parallel pol.} \\ 120\pi/\cos\theta_i; \text{ for perpendicular pol.} \end{cases}$$

$$Z_2 = \begin{cases} 40\pi \cdot \sqrt{1-\sin^2\theta_i/9}; \text{ for parallel pol.} \\ 40\pi/\sqrt{1-\sin^2\theta_i/9}; \text{ for perpendicular pol.} \end{cases}$$

Thus, the reflection coefficient for parallel polarization is

$$\Gamma_{//} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\sqrt{1-\sin^2\theta_i/9} - 3\cos\theta_i}{\sqrt{1-\sin^2\theta_i/9} + 3\cos\theta_i}$$

and the reflection coefficient for perpendicular polarization is

$$\Gamma_{\perp} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\cos\theta_i - 3\sqrt{1-\sin^2\theta_i/9}}{\cos\theta_i + 3\sqrt{1-\sin^2\theta_i/9}}$$

The results are plotted as a function of the incident angle in Figure 3.6. Obviously, the reflection coefficients are very different for different polarizations. There are some important observations:

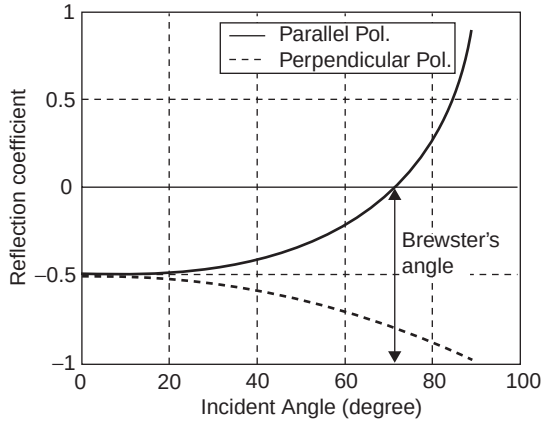


Figure 3.6 Reflection coefficient as a function of incident angle

- For parallel polarization, the reflection coefficient vanishes (= 0) at a particular incident angle, this angle is called *Brewster's angle:*

$$\theta_B = \sin^{-1}\sqrt{\frac{\varepsilon_r}{\varepsilon_r + 1}} \tag{3.32}$$

- For perpendicular polarization, the larger the incident angle, the larger the reflection coefficient (magnitude).
- At normal incidence, the reflection coefficient is the same for both polarizations.
- At 90 degrees, the reflection coefficient is 1 for parallel polarization and −1 for perpendicular polarization.

Special Case 1: when the incident angle is greater than Brewster's angle for parallel polarization, $\Gamma > 0$, as shown in Figure 3.6. Can we still employ Equation (3.30) to calculate the transmission coefficient?

The answer is no, otherwise $|T|$ would be greater than one and would fail to meet the condition set by Equation (3.31). Thus, Equation (3.30) is only valid for $\mathrm{Re}(\Gamma) < 0$. If $\mathrm{Re}(\Gamma) > 0$, Equation (3.30) should be replaced by

$$T = 1 - \Gamma = \frac{2Z_1}{Z_2 + Z_1} \tag{3.33}$$

Special Case 2: when the incident wave travels from a dense medium into a less dense medium at an angle exceeding the critical angle, what will happen?

The *critical angle* is the incident angle that gives a transmitted angle of 90 degrees when the wave is from a dense medium to a less dense medium, such as from water into air. From Snell's law, let $\sin\theta_t = 1$ to obtain this special angle

$$\theta_{ic} = \sin^{-1}\frac{\sqrt{\varepsilon_2\mu_2}}{\sqrt{\varepsilon_1\mu_1}} \tag{3.34}$$

If the relative permittivity of the water is 80, the critical angle is 6.42 degrees.

When the incident angle is greater than the critical angle, for whatever polarization, the wave will be totally internally reflected and will also be accompanied by a surface wave in the less dense medium. This surface wave decays exponentially away from the surface but propagates without loss along the surface [1]. People have developed communication systems for submarines utilizing this phenomenon.

3.3.1.1 Effects of Reflection and Transmission on Wave Polarization

Example 3.2 has clearly shown that, for non-normal incidence, the reflection coefficients are different for the two principal polarizations. As a result, if an incident wave is a combination of these two orthogonal waves, the combined signal after the reflection will be changed. For example, if the incident wave is a circular polarization, which means that the parallel component and the perpendicular component are of the same amplitudes but 90 degrees out of phase, after the reflection, the wave is no longer a circularly polarized wave but an elliptically polarized

wave – because the wave components in these two orthogonal planes are no longer the same. Similarly, if it is a linearly polarized wave, after the reflection, this linear polarization may be rotated.

The same conclusion can be drawn for the transmitted wave, since the transmission coefficients are also different for the two different polarizations if the incident angle is not zero degrees (normal incidence).

An important special case is when the reflector is a perfect conductor; here the reflection coefficient is -1. For a linearly polarized wave, this means a phase change of 180 degrees. But for a circularly polarized wave, an RCP wave becomes an LCP wave, and an LCP wave becomes an RCP after the reflection. This results in polarization mismatch. Thus, circular polarization is not recommended for indoor radio communications where plenty of reflections occur for a radio path.

3.3.1.2 Radio Waves through a Wall

A very common scenario is a radio wave passing through a wall to reach a television, mobile phone or other radio device (such as a laptop computer). How can we find out the transmission coefficient and the attenuation?

Again, we can use circuit concepts to deal with this problem, which should be much simpler than using field concepts. As shown in Figure 3.7, there are now three media. Their equivalent transmission lines have the characteristic impedances Z_1, Z_2 and Z_3, respectively. They are defined by Equation (3.28). The thickness of the wall is d. There are multiple reflections and transmissions at the boundaries. Thus, the reflected wave is now the summation of all the reflected waves, and the transmitted wave is also the summation of all the waves transmitted. They could be combined constructively (larger) or destructively (smaller). Using the transmission line model, the calculation of the reflection and transmission coefficients is much easier than the field approach (but it is still not easy!). Z_3 may be considered the load of the second transmission line. Using Equation (2.32), the input impedance at the interface between

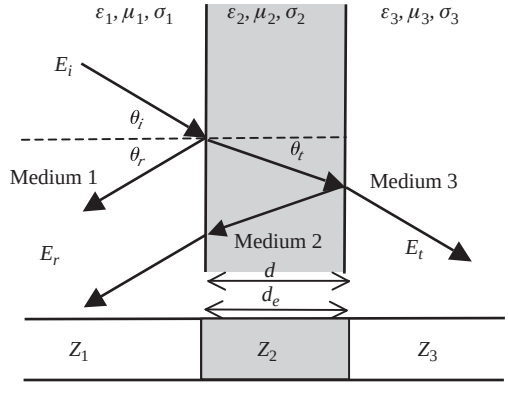


Figure 3.7 Reflection and transmission through a wall and its analogous transmission line

Medium 1 and Medium 2 can be written as

$$Z_{in} = Z_2 \frac{Z_3 + Z_2 \tanh(\gamma_2 d_e)}{Z_2 + Z_3 \tanh(\gamma_2 d_e)} \tag{3.35}$$

where $\gamma_2 = \sqrt{j\omega\mu_2(\sigma_2 + j\omega\varepsilon_2)}$ and $d_e = d/\cos\theta_t$ is the effective thickness. If the loss is negligible, we have

$$Z_{in} = Z_2 \frac{Z_3 + j Z_2 \tan(\beta_2 d_e)}{Z_2 + j Z_3 \tan(\beta_2 d_e)} \tag{3.36}$$

Thus, the reflection coefficient is:

$$\Gamma = \frac{Z_{in} - Z_1}{Z_{in} + Z_1}$$

From a power point of view, the power transmitted through the wall is

$$P_T = (P_{in} - P_R) \cdot A \tag{3.37}$$

where

P_T = power transmitted through the wall;
P_{in} = incident power;
P_R = reflected power = $P_{in} \cdot |\Gamma|^2$
A = attenuation of the wall = $e^{-2\alpha d_e}$, i.e. it is determined by the attenuation constant and the effective thickness of the wall.

Example 3.3: Reflection of a wall. A brick wall has a relative permittivity of 4 and a thickness of 20 cm, the loss is negligible.

a. If the operational frequency is 2.45 GHz for wireless applications (such as Bluetooth), plot the reflection coefficient as a function of the incident angle for both parallel and perpendicular polarizations.
b. If the incident angle is 45 degrees, plot the reflection coefficient as a function of the frequency for both parallel and perpendicular polarizations.

Solution:

We know $\varepsilon_1 = 1$, $\varepsilon_2 = 4$ and $\varepsilon_3 = 1$ and all conductivities are zero. In Medium 2, the wavelength $\lambda_2 = \lambda_0/\sqrt{4} = \lambda_0/2$, and $\beta_2 = 2\pi/\lambda_2$.

To obtain the reflection coefficient, we first use Equation (3.25) to yield

$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{\sqrt{\varepsilon_1\mu_1}}{\sqrt{\varepsilon_2\mu_2}} = \frac{1}{2}, \text{ thus } \cos\theta_t = \sqrt{1 - \sin^2\theta_t} = \sqrt{1 - \frac{1}{4}\sin^2\theta_i};$$

and then use Equation (3.28) to give

$$Z_1 = Z_3 = \begin{cases} 120\pi \cdot \cos\theta_i; \text{ for parallel pol.} \\ 120\pi/\cos\theta_i; \text{ for perpendicular pol.} \end{cases}$$

$$Z_2 = \begin{cases} 60\pi \cdot \sqrt{1-\sin^2\theta_i/4}; \text{ for parallel pol.} \\ 60\pi/\sqrt{1-\sin^2\theta_i/4}; \text{ for perpendicular pol.} \end{cases}$$

For the parallel polarization, using (3.36) yields:

$$Z_{in} = 60\pi\sqrt{1-\sin^2\theta_i/4}\ \frac{3\cos\theta_i + j\sqrt{1-\sin^2\theta_i/4}\tan(\beta_2 d/\sqrt{1-\sin^2\theta_i/4})}{\sqrt{1-\sin^2\theta_i/4} + j3\cos\theta_i\tan(\beta_2 d/\sqrt{1-\sin^2\theta_i/4})}$$

This is the load impedance to transmission line 1, thus the reflection coefficient is

$$\Gamma_{//} = \frac{Z_{in} - 120\pi \cdot \cos\theta_i}{Z_{in} + 120\pi \cdot \cos\theta_i}$$

For the perpendicular polarization, we use the same approach to obtain

$$Z_{in} = 60\pi/\sqrt{1-\sin^2\theta_i/4}\ \frac{3\sqrt{1-\sin^2\theta_i/4} + j\cos\theta_i\tan(\beta_2 d/\sqrt{1-\sin^2\theta_i/4})}{\cos\theta_i + j3\sqrt{1-\sin^2\theta_i/4}\tan(\beta_2 d/\sqrt{1-\sin^2\theta_i/4})}$$

$$\Gamma_{\perp} = \frac{Z_{in} - 120\pi/\cos\theta_i}{Z_{in} + 120\pi/\cos\theta_i}$$

a. When the incident angle is set as the variable, the reflection coefficients for both polarizations are as plotted in Figure 3.8. The important observations are:
   - the reflection coefficients are about the same at small incident angles, and at 90 degrees;
   - there are two troughs for parallel polarizations but one for perpendicular polarization;

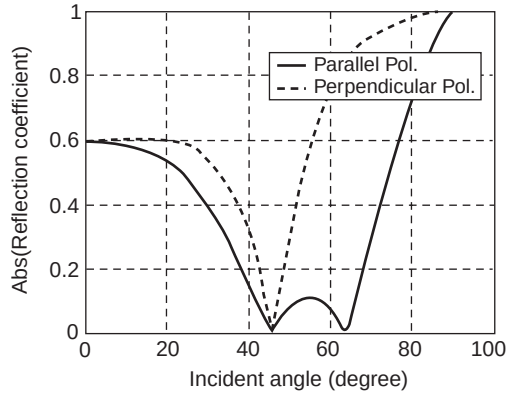


Figure 3.8 Reflection coefficient of a wall as a function of the incident angle

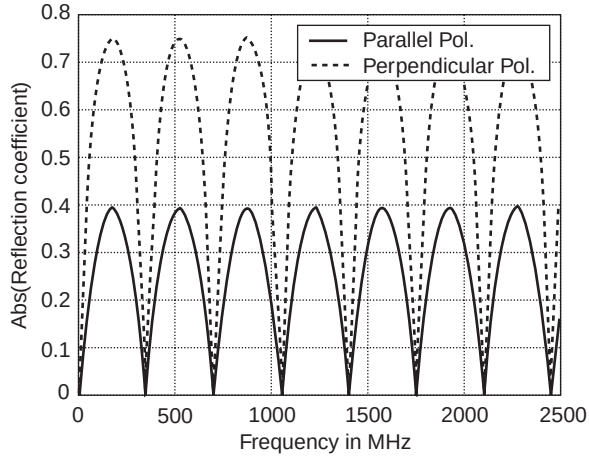


Figure 3.9 The reflection coefficient of a wall as a function of the frequency

- the reflection coefficient of perpendicular polarization is always greater than (or the same as) that of parallel polarization.

b. When the frequency is set as the variable, the reflection coefficients for both polarizations are as plotted in Figure 3.9. We can see that:
   - The reflection coefficients are periodic functions of the frequency. The period for both polarizations is the same, 350 MHz in this particular case (for the normal incident, the period is 375 MHz), i.e. its wavelength is half of the effective thickness. Thus, *the reflection coefficient is minimized when the thickness of the wall is an integer of half of the effective wavelength*. This important conclusion can be used for antenna radome and housing design.
   - *The reflection reaches the maximum when* $d_e = \lambda/4 + n\lambda/2$, $n = 0, 1, 2\ldots$
   - The reflection coefficients are very small for low frequencies. This is why low-frequency signals can easily penetrate buildings.
   - Again, the reflection coefficient of perpendicular polarization is always greater than (or the same as) that of parallel polarization.

In this example, since the conductivity is zero, the waves which are not reflected by the wall will pass through the wall without attenuation and the transmission coefficient can be calculated using the reflection–transmission equation (3.31).

It is clear now that radio wave reflection and transmission are complicated. The coefficients are functions of frequency, incident angle and polarization, as well as the dielectric properties of the media.

3.3.2 Diffraction and Huygens's Principle

Diffraction is the apparent bending and spreading of waves when they meet an obstacle. It can occur with any type of wave, including sound waves, water waves and electromagnetic waves. As a simple example of diffraction, if you speak into one end of a cardboard tube, the sound

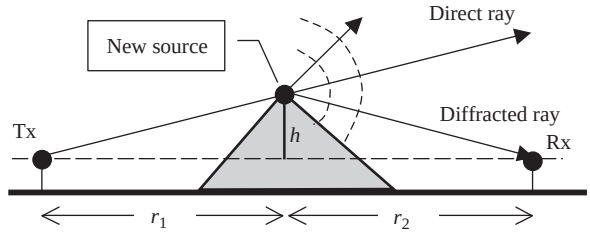


Figure 3.10 Radio wave diffraction over a knife-edge obstacle

waves emerging from the other end spread out in all directions, rather than propagating in a straight line like a stream of water from a garden hose.

The foundation of diffraction theory is based on *Huygens's Principle*, which states that *each point on a primary wave front can be considered a new source of a secondary spherical wave*, as suggested in Figure 3.10. The new source is considered an equivalent source, which can be expressed mathematically as

$$\begin{aligned} \boldsymbol{J_S} &= \hat{\boldsymbol{n}} \times \boldsymbol{H} \\ \boldsymbol{M_S} &= -\hat{\boldsymbol{n}} \times \boldsymbol{E} \end{aligned} \tag{3.38}$$

where $\boldsymbol{J}_S$ and $\boldsymbol{M}_S$ are the equivalent surface electric current and magnetic current on an imaginary surface S, respectively; $\hat{\boldsymbol{n}}$ is the unit vector, normal to the surface S. Thus, the original problem can be replaced by an equivalent source which produces the same electric and magnetic fields outside the (enclosed) surface S. Thus, Huygens's principle is closely linked to Love's *equivalence principle*, which was developed by replacing an actual radiating source by an equivalent source. This principle is very useful for analyzing aperture-type and slot antennas, which will be discussed in Chapter 5.

When the distance from the obstacle to the receiver (Rx) is much larger than the height of the obstacle, which could be a mountain or a building, the relative (to the direct ray) power density S can be approximated by [1]

$$S = \frac{\lambda}{4\pi^2 h^2}\left(\frac{r_1 r_2}{r_1 + r_2}\right) \tag{3.39}$$

where λ is the wavelength. r_1 and r_2 are the distances to the transmitter and receiver, respectively and h is the height of the obstacle above the reference line.

Thus, diffraction by obstacles is an important mechanism for radio propagation, especially when there is no line-of-sight path available. Radio broadcasting signals can pass over hills/mountains and mobile radio signals can reach to the street; these are just some of the examples benefitting from radio wave diffraction. Radar is able to detect various targets, which is also partially due to diffraction of the radar signal from the target.

3.3.3 Scattering

Unlike the other propagation mechanisms where the size of the medium or the obstacle is much larger than the wavelength, *scattering* occurs when the obstacle is comparable to or even smaller

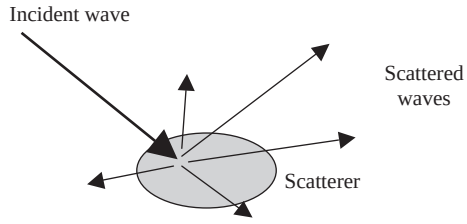


Figure 3.11 Radio wave scattering

than the wavelength. It is the process by which small particles/obstacles in a medium of different dielectric properties diffuse a portion of the incident wave in all directions. In scattering, there are no energy transformation results, only a change in the spatial distribution of the radiation, as shown in Figure 3.11. Back scattering has been employed for radar applications, such as target detection and weather forecasting.

Along with absorption, scattering is a major cause of the attenuation of radio propagation by the atmosphere. Scattering varies as a function of the ratio of the particle/obstacle diameter to the wavelength of the wave. When this ratio is less than about one-tenth, *Rayleigh scattering* occurs, in which the scattering coefficient varies inversely as the fourth power of the wavelength – this result can be used to explain why the sky is red at sunrise and sunset but blue at midday (the wavelength of red light is long and the wavelength of blue light is short, thus most transmitted light is red but most scattered light is blue). At larger values of the ratio of the particle diameter to the wavelength, the scattering varies in a complex fashion described by the *Mie theory*; at a ratio of the order of 10, the laws of *geometric optics* (where the wave can be treated as a ray) begin to apply.

3.4 Radio Wave Propagation Characteristics in Media

Radio waves propagating through a radio channel may undergo reflection, transmission or refraction, diffraction and scattering. Attenuation or absorption is another important aspect that we have not yet discussed properly. The same radio wave propagating through different media may exhibit very different features. In this section we are going to briefly examine radio wave propagation characteristics in some common media.

3.4.1 Media Classification and Attenuation

From an electromagnetics point of view, materials can be classified as conductive, semiconductive or dielectric media. The electromagnetic properties of materials are normally functions of the frequency, so are the propagation characteristics. Recall that, in Chapter 1, we introduced the complex permittivity and defined the loss tangent as the ratio of the imaginary to the real parts of the permittivity, which is Equation (1.36):

$$\tan\delta = \frac{\varepsilon''}{\varepsilon'} = \frac{\sigma}{\omega\varepsilon}$$

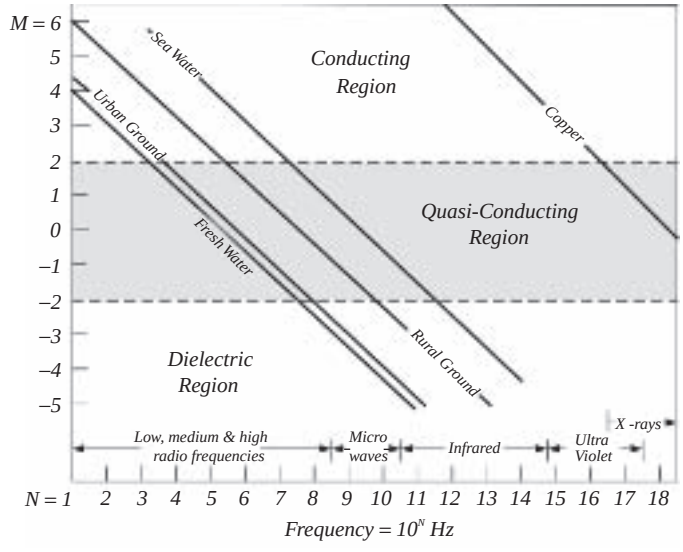


Figure 3.12 Classification of media as a function of frequency ($\tan\delta = 10^M$). (J. D. Kraus and D. A. Fleisch, Electromagnetics with Application, 5th edition, McGraw-Hill, 1999. Reproduced by permission of The McGraw-Hill Companies)

This has taken the frequency as well as the normal permittivity and conductivity of the medium into account. The specific classifications are given in [1] as

- **Conductor:** $\tan\delta = \dfrac{\sigma}{\omega\varepsilon} > 100$
- **Semi-conductor:** $0.01 < \tan\delta = \dfrac{\sigma}{\omega\varepsilon} < 100$
- **Dielectric:** $\tan\delta = \dfrac{\sigma}{\omega\varepsilon} < 0.01$

The classification of some common media as a function of frequency is shown in Figure 3.12, where M is determined by $\tan\delta = 10^M$. It is important to note that:

- The medium classification is indeed frequency dependent.
- Most materials are in the conducting region at low frequencies ($<$ 10 kHz) but in the dielectric region at high frequencies ($>$ 1 GHz). The ground and water are just two examples.
- Figure 3.12 is just an approximation to illustrate material properties over a wide frequency range. Real-world materials are more complicated. A typical complex dielectric permittivity spectrum is shown in Figure 3.13. A Cole–Cole model or Debye relaxation model is often used to approximate this frequency response. The permittivity ε and conductivity σ are not constant but functions of frequency as well.

This classification is useful for evaluating the EM properties of a medium in terms of the loss tangent but is not accurate for classifying whether a medium is lossy or not. Recall that

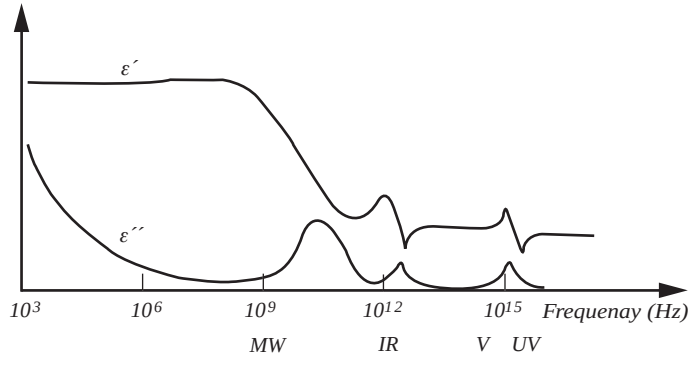


Figure 3.13 A typical complex permittivity spectrum

the attenuation constant (it is, in fact, not a constant) is given by (3.4) as

$$\alpha = \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\left(\sqrt{1+\frac{\sigma^2}{\varepsilon^2\omega^2}}-1\right)\right]^{1/2} = \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\left(\sqrt{1+\tan^2\delta}-1\right)\right]^{1/2} \tag{3.40}$$

The loss tangent is just a term in the bracket. The attenuation constant is actually proportional to the frequency if the loss tangent is fixed. Generally speaking, the dominant feature of radio wave propagation in media is that the attenuation increases with the frequency – this seems to contradict the material classification. Figure 3.12 is not a good reference for propagation loss characteristics, but a useful plot of the loss tangent against the frequency.

It is now clear that many common media may be considered to be conducting materials at low frequencies and dielectric materials at high frequencies, but their attenuation increases with frequency. This is why, generally speaking, low frequencies have been used for longer distance communications and high frequencies are employed for shorter distance communications.

3.4.1.1 Propagation through the Ionosphere

It is important to take into account the propagation characteristics of the ionosphere when long-distance and satellite-type communications are considered. The *ionosphere* is the region above the troposphere (where the air is), from about 80 to 400 km above the Earth. It is a collection of ions, which are atoms that have some of their electrons stripped off, leaving two or more electrically charged objects. The sun's rays cause the ions to form, and these then slowly recombine.

The propagation of radio waves in the presence of ions is drastically different from that in air, which is why the ionosphere plays an important role in most modes of propagation. The major effects of the ionosphere on radio waves are:

- Reflection at low frequencies (up to about 30 MHz). Thus, the ionosphere and the Earth can form a kind of waveguide, letting a wave propagate over a very long distance using ground/surface mode.

- Scattering, refraction and absorption when high-frequency waves (above 100 MHz) pass through it.
- Faraday rotation: the wave polarization plane/line is rotated through the ionosphere. The amount of rotation is dependent on the thickness and charge density of the ionosphere, which are functions of time (and where the sun is), and is also dependent on the frequency. The rotation is small at very high frequencies (> 10 GHz), which is why linear polarizations may be employed for satellite communications or broadcasting if the frequency is high enough.

3.4.1.2 Propagation in Rain

Rain is an undesired medium of a radio channel. It causes a considerable number of problems every year for radio communications and radar systems. The major effect of rain on radio waves is attenuation due to absorption and scattering over a wide range of the spectrum. The attenuation depends on a number of things and it can be represented by [2]

$$A = aR^b, \text{ dB/km} \tag{3.41}$$

where R is the rainfall rate in mm/h, and a and b are constants that depend on frequency and temperature of the rain. The temperature dependence is due to the variation of dielectric permittivity of water with temperature.

A light drizzle corresponds to a rainfall rate of 0.25 mm/h, light rain to 1 mm/h, moderate rain to 4 mm/h, heavy rain to 16 mm/h and cloud bursts up to many cm/h. Figure 3.14 is an illustration of how the attenuation is linked to the rainfall rate and frequency [2].

The figure shows clearly that the attenuation around 10 GHz is much smaller than 30 and 100 GHz. At a rainfall rate of 10 mm/h, the attenuation is increased from 0.2 dB/km at 10 GHz to about 2 dB/km at 30 GHz (ten times more). If we take the sky noise temperature into account, a low-noise and low-loss radio 'window' exists between 1 and about 15 GHz – this is why most satellite communication systems operate within this frequency band.

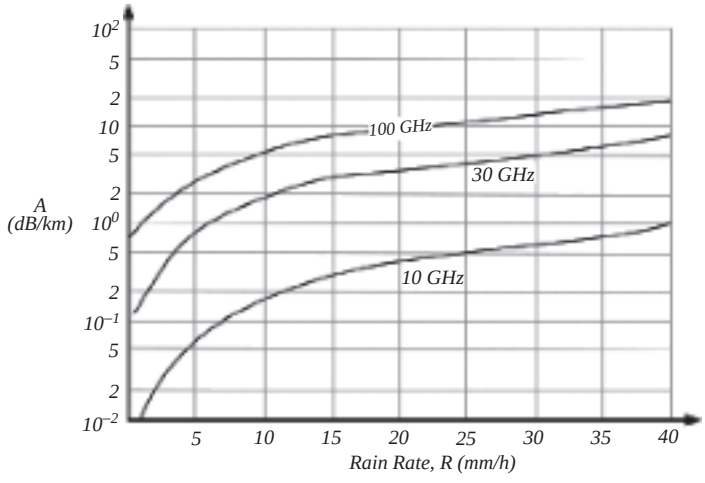


Figure 3.14 Rain attenuation as a function of rainfall rate and frequency. (R. E. Collin, Antennas and Radiowave Propagation, McGraw-Hill, Inc., 1985. Reproduced by permission of The McGraw-Hill Companies)

3.4.1.3 Propagation in Snow

Some parts of the world have no snow at all, but other parts have to face the problem most of the time. The attenuation in dry snow is an order of magnitude less than that in rain for the same precipitation rate. But attenuation by wet snow is comparable to that in rain and may even exceed that of rain at millimeter wavelengths. It is difficult to specify the attenuation in any simple form.

3.4.1.4 Propagation through Fog

Fog is another unfriendly medium we have to deal with. The attenuation equation is the same as that for rain but with much smaller attenuation.

3.5 Radio Wave Propagation Models

For radio communication system designers, it is important to be able to predict the radio wave propagation *pathloss*, which is defined as the difference between the power transmitted and the power received at the destination. Using the pathloss information, the designer is able to optimize the system and ensure sufficient radio coverage. That is, to establish the link budget and figure out where to place the antenna and what the required transmitted power is to cover the desired area. Since this is a very important subject, it has been studied comprehensively by many researchers over the years and some propagation models have been developed for various scenarios. Here we are going to discuss some basic propagation models closely linked to antennas.

3.5.1 Free Space Model

If there is just one ray between the transmitting antenna and the receiving antenna (the line-of-sight case), such as in satellite/space communications, as shown in Figure 3.15, the received power can be obtained using the well-known *Friis transmission formula*, which will be discussed in Chapter 4, that is

$$P_r = P_t \left(\frac{\lambda}{4\pi r}\right)^2 G_t G_r \tag{3.42}$$

where P_r is the received power, P_t is the transmitted power, G_t and G_r are the gains of the transmitting and receiving antennas, respectively and r is the distance between the transmitting and receiving antennas. Without considering the antenna performance (let $G_t = G_r = 1$), the pathloss L_P can be found from Equation (3.42) as

$$L_P = 10\log_{10}\left(\frac{P_t}{P_r}\right) = 20\log_{10} f + 20\log_{10} r - 147.6 \text{ (dB)} \tag{3.43}$$

where f is the frequency. This means that the pathloss is proportional to the frequency squared and the distance squared. The larger the distance, the larger the pathloss; the higher the frequency, the larger the pathloss. For a fixed distance or frequency, the pathloss is 20 dB/decade.

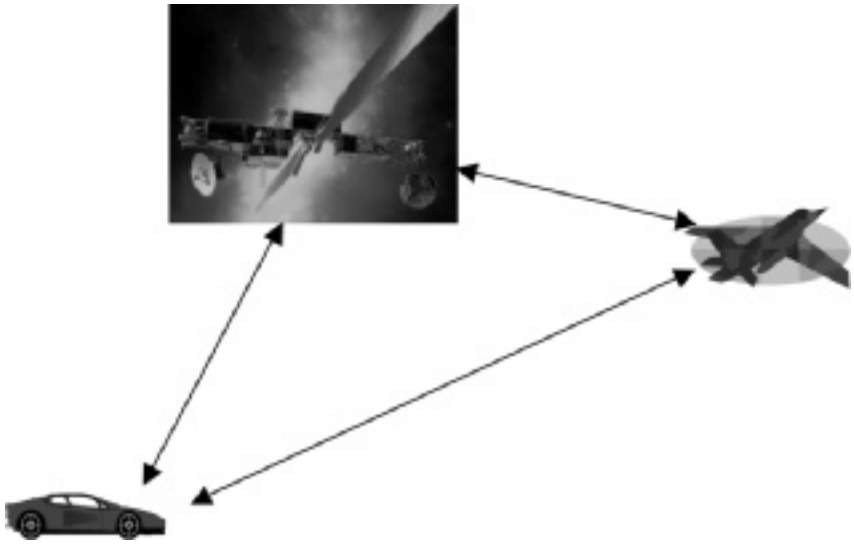


Figure 3.15 Free space communications

3.5.2 Two-ray Model/Plane Earth Model

As shown in Figure 3.16, the two-ray model can be applied to many terrestrial communications scenarios, such as radio broadcasting and radio communications in rural environments. The transmitted signals reach the destination via the line-of-sight path and the path reflected by the ground. The result becomes complicated because the signals may be combined constructively or destructively, depending on the reflection coefficient of the ground and the phase difference of these two rays.

The path difference between these two rays is (using the image theory to be discussed in the next chapter and assuming $d \gg h_1 + h_2$)

$$\Delta r = \sqrt{(h_1 + h_2)^2 + d^2} - \sqrt{(h_1 - h_2)^2 + d^2} \approx \frac{2h_1 h_2}{d}$$

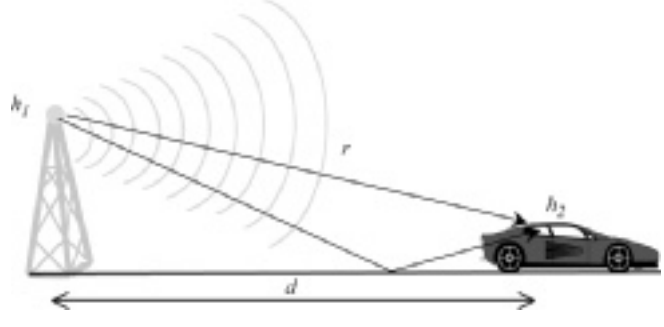


Figure 3.16 Two-ray model

When this difference is half of the wavelength, the phase difference is 180 degrees (out of phase), the separation is

$$d_f = \frac{4h_1h_2}{\lambda} \tag{3.44}$$

and is called the *first Fresnel zone distance*. It is useful to note the following results:

- when the separation of the two antennas is smaller than this distance, $d < d_f$, the pathloss is about 20 dB/decade;
- when the separation is greater than this distance, $d > d_f$, the pathloss is 40 dB/decade, because it can be shown [3] that, in this case, the received power

$$P_r = P_t \left(\frac{h_1h_2}{d^2}\right)^2 \tag{3.45}$$

Hence, the pathloss expressed in dB is

$$L_P = -10\log_{10}\left(\frac{h_1h_2}{d^2}\right)^2 = 40\log_{10}(d) - 20\log_{10}(h_1h_2),\ \text{dB} \tag{3.46}$$

It is apparent that the higher the base-station antenna and the mobile antenna, the smaller the pathloss. Thus, there is little wonder that all mobile operators want to erect their antennas as high as possible!

3.5.3 Multipath Models

For the multipath case (more than two rays), no analytical equations have been obtained to give an accurate prediction of the radio propagation pathloss. But empirical and statistical representations are available for various scenarios.

Most of the popular outdoor pathloss prediction tools are based on Okumura and Hata's formulation, which was based on a huge amount of measured data for frequencies between 100 MHz and 3 GHz [3].

For indoor scenarios, many researchers have shown that indoor pathloss actually obeys the distance power law as

$$L_P(d) = L_P(d_0) + 10n\log\left(\frac{d}{d_0}\right) + X,\ \text{dB} \tag{3.47}$$

where d_0 is the reference distance (normally it is 1 m), the power index n depends on the frequency, surroundings and building type and X represents a normal random variable in dB having a standard deviation of σ dB. Typical values for various buildings are provided in Table 3.1. It is apparent that the value n is normally (but not always) larger than 2, which is the free space case. It also shows that the higher the frequency, the larger n for the same scenario.

Table 3.1 Path-loss exponent and standard deviation measured in different buildings

Building	Frequency (MHz)	Index n	Standard deviation σ (dB)
Retail Stores	914	2.2	8.7
Office (hard partition)	1500	3.0	7.0
Office (soft partition)	1900	2.6	14.1
Office (soft partition)	915	2.0–2.8	9–14
Open-plan factory	1300	2.0–2.4	3.7–9.0
Suburban home	900	3.0	7.0
Factory LOS	1300	1.6–2.0	3.0–7.0
Factory OBS	1300	3.3	6.8

For a multipath environment, in addition to the pathloss, there are many other important characteristics, which include the following:

- **Multipath fast fading:** the received signal changes significantly (> 30 dB) over a very short distance (few wavelengths), resulting from the complex and vector summation of signals.
- **Delay spread:** multiple copies of the original signal arrive at the destination at different times through different paths, which may cause dispersion and *inter-symbol interference, ISI*, (the equalizer is a device developed to remove all delayed waves and combat the ISI problem). The RMS (root-mean-square) delay spread T_{RMS} (in seconds) is used to characterize this feature. The delay spread is often employed to define the channel's *coherence bandwidth* B_C (similar to a filter's bandwidth – a radio channel can be considered a filter in a certain sense). They are linked by the following equation:

$$B_C = \frac{1}{2\pi T_{RMS}}, (Hz) \tag{3.48}$$

If the channel bandwidth is greater than the message bandwidth, all the frequency components in the message will arrive at the receiver with little or no distortion and the ISI can be negligible. The channel is described as *flat fading*.

If the channel bandwidth is smaller than the message bandwidth, all the frequency components in the message will arrive at the receiver with distortion and ISI. In this case the channel is described as *frequency selective fading*.

Statistically, the power density function (PDF) of the received (short-term) signal envelope follows a certain distribution, depending on the specific propagation channel. According to the PDF distribution function, the channel can be classified as one of the following:

- when there is a line-of-sight ray, it follows a Gaussian distribution and this channel is therefore called the *Gaussian channel*;

- When there is a partial line-of-sight ray (the path is partially blocked by obstacles such as trees), it follows a Rician distribution and this channel is therefore called the *Rician channel*;
- when there is no line-of-sight ray, it follows a Rayleigh distribution and this channel is therefore called the *Rayleigh channel*.

For the same signal-to-noise ratio (SNR), a communication system can obtain the best bit error rate (BER) from a Gaussian channel and the worst BER from a Rayleigh channel.

Antenna diversity techniques and MIMO antennas, which will be discussed later, were developed specifically for multipath environments.

3.6 Comparison of Circuit Concepts and Field Concepts

We have now introduced both circuit concepts and field concepts, which may seem to be completely different. However, if we compare these two sets of parameters, some interesting correspondence can be obtained. For example, the product of the voltage and current is power in circuit concepts, whilst the cross product of the electric field and magnetic field is the power density in field concepts; the ratio of the voltage to the current is the impedance, whilst the ratio of the electric field to the magnetic field is the intrinsic impedance. Let's take another look at how the electric field and magnetic field are distributed around a two-wire transmission line, as shown in Figure 3.17. It is not difficult to conclude that *the fundamental correspondences of these two sets of concepts are*:

- the voltage V corresponds to the electric field $\boldsymbol{E}$ via $V = C_1 \int_{Line1}^{Line2} \boldsymbol{E} \bullet \boldsymbol{ds}$
- the current I corresponds to the magnetic field $\boldsymbol{H}$ via $I = C_2 \oint \boldsymbol{H} \bullet \boldsymbol{ds}$

C_1 and C_2 are just two constants. A list of the correspondences between circuit and field concepts is presented in Table 3.2.

3.6.1 Skin Depth

Now let us take the skin depth as an example to show the difference between circuit concepts and field concepts in applications.

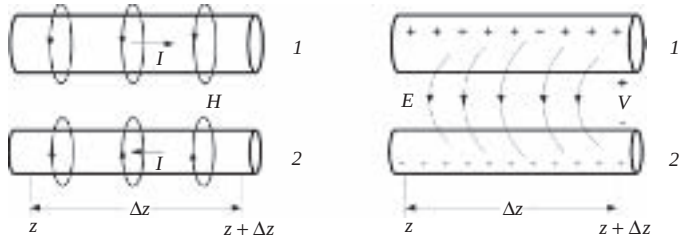


Figure 3.17 The linkage between circuit concepts and field concepts

Table 3.2 Correspondences between circuit concepts and field concepts

	Circuit		Field	
Complex voltage	$V(t, z)$	in V	$\boldsymbol{E}(t, z)$	in V/m
Complex current	$I(t, z)$	in A	$\boldsymbol{H}(t, z)$	in A/m, or $\boldsymbol{J}(t, z)$
Complex power flow	VI^*	in W	$\boldsymbol{E} \times \boldsymbol{H}^*$	in W/m^2
Impedance	Z	in Ω	η	in Ω
Admittance	$1/R$	in S	σ	in S/m
Current	V/R	in A	$\mathbf{J} = \sigma\mathbf{E}$	in A/m^2

3.6.1.1 Skin Depth – Field Concepts

From the field concept point of view, a radio wave traveling into a lossy medium can be described by Equation (3.7). That is

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_0 e^{-\alpha z + j(\omega t - \beta z)}$$

The *skin depth* is defined as the distance δ through which the amplitude of a traveling plane wave decreases by factor $1/e$ and is therefore

$$\delta = \frac{1}{\alpha} \tag{3.49}$$

For a good conductor ($\frac{\sigma}{\omega\varepsilon} > 100$), the attenuation constant can be approximated as

$$\alpha = \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\left(\sqrt{1 + \frac{\sigma^2}{\varepsilon^2\omega^2}} - 1\right)\right]^{1/2} \approx \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\frac{\sigma}{\varepsilon\omega}\right]^{1/2} = \sqrt{\frac{\omega\mu\sigma}{2}}$$

The skin depth can therefore be expressed as

$$\delta \approx \sqrt{\frac{2}{\omega\mu\sigma}} = \sqrt{\frac{1}{\pi f \mu\sigma}} \tag{3.50}$$

This means that

- the higher the frequency, the smaller the skin depth;
- the larger the permeability, the smaller the skin depth;
- the larger the conductivity, the smaller the skin depth.

For a good conductor, the permittivity has little effect on the skin depth.

It should be highlighted that the wave amplitude is reduced by a factor of $1/e$, or 37%, or 8.686 dB over one skin depth.

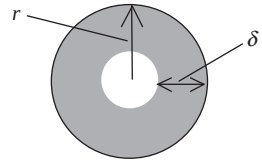


Figure 3.18 A conducting wire of skin depth δ

3.6.1.2 Skin Depth – Circuit Concepts

From the circuit concept point of view, the current density $\boldsymbol{J}$ in an infinitely thick plane conductor decreases exponentially with depth from the surface as follows:

$$J = J_0 e^{-z/\delta}$$

where δ is a constant called the *skin depth*. This is defined as the depth below the surface of the conductor at which the current density decays to $1/e$ (about 37%) of the current density at the surface. Mathematically it is the same as Equation (3.50). As a result, when calculating the per-unit-length resistance of a conducting wire,

$$R = \frac{1}{S\sigma}$$

where S is now the effective area rather than the whole cross-section of the wire when the wire radius $r > \delta$. In Figure 3.18, this effective area can be approximated as $S = \pi r^2 - \pi(r-\delta)^2 \approx 2\pi r\delta$ (not $S = \pi r^2$). Thus, the per-unit-length resistance is now

$$R \approx \frac{1}{2\pi r\delta\sigma} = \frac{\sqrt{f\mu}}{2r\sqrt{\pi\sigma}} \tag{3.51}$$

It has become apparent that the skin effect causes the effective resistance of the conductor to increase with the frequency of the current. To illustrate the effects of skin depth, the skin depth and resistance of a 3 cm gold rectangular track of a device (as part of a microstrip line or CPW) are given in Table 3.3 (conductivity $\sigma = 4.1 \times 10^7$ S/m). The width and thickness of

Table 3.3 Skin depth and resistance of a gold track of dimensions 7 μm × 16 μm × 30,000 μm

Frequency (MHz)	Skin depth (μm)	Resistance (ohms)
1	78.60	6.53
1000	2.49	8.16
2500	1.57	11.72
10,000	0.79	21.72
20,000	0.56	30.07
40,000	0.39	41.91

the track are 16 μm and 7 μm, respectively. The skin depth is small and the resistance changes significantly with the frequency.

The effects of skin depth are not just limited to the resistance; it may cause other problems such as overheating in certain areas of the circuit or antenna, which should also be taken into account in system design.

3.7 Summary

In this chapter we have obtained the wave solutions and introduced a set of concepts and parameters for field analysis, which include the plane wave, polarization and intrinsic impedance. Careful examinations of radio wave reflection, transmission, diffraction and scattering have been undertaken, and the focus has been on the analysis of radio reflection and transmission, the two most common propagation mechanisms. Radio propagation models for various scenarios and channel characteristics have been introduced and discussed. A comparison of field concepts and circuit concepts has revealed the interesting correspondence between these two sets of concepts.

References

[1] J. D. Kraus and D. A. Fleisch, *Electromagnetics with Application*, 5th edition, McGraw-Hill, 1999.
[2] R. E. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, Inc., 1985.
[3] T. S. Rappaport, *Wireless Communications*, 2nd edition, Prentice Hall, 2002.

Problems

Q3.1 Explain the concept of a plane wave with the aid of a diagram.

Q3.2 Compared with other forms of waves (such as sound waves), what is the unique feature of EM waves?

Q3.3 The radio propagation constant is given by Equation (3.3). Prove that the attenuation and phase constants can be expressed as

$$\alpha = \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\left(\sqrt{1+\frac{\sigma^2}{\varepsilon^2\omega^2}}-1\right)\right]^{1/2}$$

$$\beta = \omega\sqrt{\mu\varepsilon}\left[\frac{1}{2}\left(\sqrt{1+\frac{\sigma^2}{\varepsilon^2\omega^2}}+1\right)\right]^{1/2}$$

respectively.

Q3.4 For a lossy/conducting medium, $\tan\delta = \frac{\sigma}{\omega\varepsilon} > 100$. Find simplified expressions for the attenuation and phase constants. If this medium is sea water, find its attenuation and phase constants, phase velocity and wavelength in the medium at 1 MHz. Make a comparison with their free space counterparts.

Q3.5 If the electric field of a wave can be expressed as

$$\boldsymbol{E} = 2e^{j(\omega t+\beta z)}\hat{\boldsymbol{x}}$$

where t is the time and z is the distance along the z-axis.

a. What are ω and β? The propagation direction? The polarization?
b. Verify that this is a solution of the wave equation.
c. Obtain the wave velocity in free space and the magnetic field $\boldsymbol{H}$.
d. Find the power flow density of the wave.
e. If this wave is reflected by a perfect conductor at $z = 0$, obtain the reflection coefficient and write down the expression of the reflected wave.

Q3.6 Explain the concept of circular polarization. Give an example of a right-hand circular polarization and suggest a typical application for such a signal.

Q3.7 What is Brewster's angle? What is the critical angle?

Q3.8 Explain the concept of skin depth. If a copper box is employed to host a sensitive high data rate digital circuit and the minimum electric field attenuation required at 1 GHz is 100 dB, find the minimum thickness of the box. This is equivalent to how many skin depths?

Q3.9 An important application of a $\lambda/4$ matching plate is for eliminating reflections. Prove that a wave is matched through a thick slab of $\varepsilon_r = 4$ dielectric by means of two $\lambda/4$ plates of $\varepsilon_r = 2$ on each side of the slab.

Q3.10 If a plane wave at 1 GHz propagates normally to a ground plane having constants: relative permittivity $\varepsilon_r = 9$, conductivity $\sigma = 0$ and relative permeability $\mu_r = 1$, find

a. what the intrinsic impedance of the medium is;
b. how much energy (in % and dB) is reflected.

If a layer of paint is used to improve the matching, what are its desired characteristics, such as ε_r and thickness?

If a layer of paint is used to achieve the maximum reflection, what are its desired characteristics?

Q3.11 A right-hand circularly polarized plane wave is reflected by a good conductor. What is the polarization of the reflected wave? If this wave is reflected by a concrete ground plane, comment on the polarization of the reflected wave again and justify your comments.

Q3.12 Explain the concept of scattering. Why is the sky red at sunrise (and blue at noon)?

Q3.13 A mobile radio wave at 1800 MHz of linear polarization is reflected by a 20 cm brick wall with a relative dielectric constant of 4.5 and conductivity of 0. If the incident angle is 30 degrees,

a. calculate the reflection and transmission coefficients if the wave has perpendicular polarization;
b. calculate the reflection and transmission coefficients if the wave has parallel polarization;
c. if the incident wave has neither perpendicular nor parallel polarization, how can we calculate the reflection coefficient?

Q3.14 Discuss the effects of fog on radio propagation and explain how to obtain the attenuation factor using a mathematical expression. Illustrate how the attenuation changes against frequency with the aid of a diagram.

Q3.15 The indoor radio propagation channel is of interest to many wireless engineers. Pathloss and delay spread are two of the most important parameters of a radio channel. Explain the concepts of these two parameters.

Q3.16 Radio propagation models are very important tools for radio system designers and planners. Derive the pathloss for the free space model and two-ray model, respectively, and then compare them in terms of the variation against the frequency and distance.

Q3.17 Radio wave absorbing materials are widely used for antenna measurements. A lossy mixture of a high-μ (ferrite) and a high-ε (barium titanate) material can be used effectively for wave absorption with the ratio μ/ε equal to that for free space. Let a 1 GHz plane wave be incident normally on a solid ferrite-titanate slab of thickness $d = 2$ cm and $\mu_r = \varepsilon_r = 30(1 - j1)$. The medium is backed by a flat conducting sheet. How much is the reflected wave attenuated with respect to the incident wave? Express this in dB.

4

Antenna Basics

We have introduced circuit concepts and field concepts and studied transmission lines and radio waves in previous chapters. In this chapter we are going to study antenna theory, see how antennas are linked to radio waves and transmission lines and introduce the essential and important parameters of an antenna from both the circuit point of view and the field point of view. It is hoped that, through this chapter, you will become familiar with the antenna language and gain a better understanding of antennas. At the end, you will know how they work and what the design parameters and considerations are.

4.1 Antennas to Radio Waves

In Chapter 2 we introduced Maxwell's equations, which reveal the fundamental relations between the electric field, the magnetic field and the sources. Again, we only discuss a single-frequency source case (an arbitrary case can be considered to be the combination of many single-frequency sources), thus Maxwell's equations can be written as

$$\begin{aligned} \nabla \times \boldsymbol{E} &= -j\omega\mu\boldsymbol{H} \\ \nabla \times \boldsymbol{H} &= \boldsymbol{J} + j\omega\varepsilon\boldsymbol{E} \\ \nabla \bullet \boldsymbol{E} &= \rho/\varepsilon \\ \nabla \bullet \boldsymbol{H} &= 0 \end{aligned} \tag{4.1}$$

Using a similar process to that used to derive Equation (3.2), we can obtain

$$\nabla^2\boldsymbol{E} + \omega^2\mu\varepsilon\boldsymbol{E} = j\omega\mu\boldsymbol{J} + \nabla(\rho/\varepsilon) \tag{4.2}$$

This is an equation which links the radiated electric field (no magnetic field) directly to the source. To solve this equation, boundary conditions are required. For an open boundary, which means that the field vanishes when the distance from the source V to the field point becomes infinite, as shown in Figure 4.1, the solution of Equation (4.2) in a uniformed medium

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

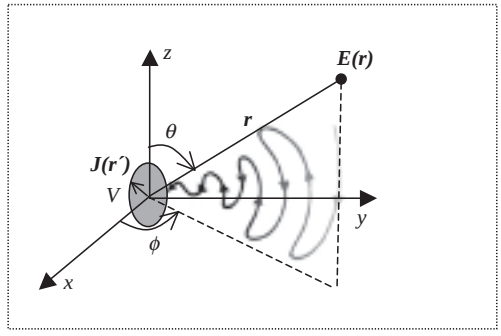


Figure 4.1 Coordinates and radio waves generated by a time-varying source

(μ and ε) is [1]

$$\boldsymbol{E}(\boldsymbol{r}) = -j\omega\mu \int_V \boldsymbol{J}(r') \frac{e^{-j\beta\cdot|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi\,|\boldsymbol{r}-\boldsymbol{r}'|} dv' + \frac{1}{j\omega\varepsilon} \nabla \left(\nabla \bullet \int_V \boldsymbol{J}(r') \frac{e^{-j\beta\cdot|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi\,|\boldsymbol{r}-\boldsymbol{r}'|} dv' \right) \tag{4.3}$$

where $\boldsymbol{r}$ is the distance vector from the origin to the observation point and $\boldsymbol{r}'$ is from the origin to the source point. *This equation gives the radiated electric field from a time-varying current* $\boldsymbol{J}$ *(the time factor* $e^{j\omega t}$ *is omitted here) and is the very foundation of antenna theory – it reveals how the antenna is related to radio waves.* Only a time-varying current (vibrating charges as shown in Figure 4.1) can generate a radio wave – not a DC current or static charges. *Antenna design is all about how to control the current distribution* $\boldsymbol{J}$ *and hence to obtain the desired radiated field* $\boldsymbol{E}$. Antenna theory may be summarized by this single but complex equation, which includes vector partial differentiation and integration. It is normally not possible to yield an analytical expression of the radiated field.

4.1.1 Near Field and Far Field

However, the analytical solution is obtainable for some very simple cases. For example, when the source is an *ideal current element* with length Δl and current value I, the current density vector can be expressed as

$$\boldsymbol{J} = \hat{\boldsymbol{z}} I\,\Delta l$$

This current is electrically short (i.e. $\Delta l \ll \lambda$) and fictitious, but very useful for antenna analysis. Replacing $\boldsymbol{J}$ with this representation in Equation (4.3), the radiated electric field can be found as

$$\begin{aligned} E_r &= 2\frac{I\,\Delta l}{4\pi}\eta\beta^2\cos\theta\left(\frac{1}{\beta^2 r^2} - \frac{j}{\beta^3 r^3}\right)e^{-j\beta r} \\ E_\theta &= \frac{I\,\Delta l}{4\pi}\eta\beta^2\sin\theta\left(\frac{j}{\beta r} + \frac{1}{\beta^2 r^2} - \frac{j}{\beta^3 r^3}\right)e^{-j\beta r} \\ E_\phi &= 0 \end{aligned} \tag{4.4}$$

where η is the intrinsic impedance of the medium. We can use this and the first equation in Equation (4.1) to yield the magnetic field:

$$H_r = 0;\ H_\theta = 0$$
$$H_\phi = \frac{I\Delta l}{4\pi}\beta^2 \sin\theta \left(\frac{j}{\beta r} + \frac{1}{\beta^2 r^2}\right) e^{-j\beta r} \tag{4.5}$$

It is apparent that, for this simple case, the electric field has E_θ and E_r components, whilst the magnetic field has just an H_ϕ component. The other field components are zero.

When the angle θ is 90 degrees, the magnetic field reaches the maximum. The electric field, magnetic field and their ratio are shown in Figure 4.2 as a function of βr for a given frequency. It is interesting to note that

- the electric field is always greater than the magnetic field;
- when $\beta r > 1$, the ratio of $|\mathbf{E}/\mathbf{H}| = \eta$ is about 377 Ω in free space;
- when $\beta r < 1$, as the distance increases, the electric field reduces at a much faster rate (60 dB/decade) than the magnetic field (40 dB/decade).

When the distance is fixed, the electric field, magnetic field and their ratio are shown in Figure 4.3 as a function of β (which is proportional to the frequency). It is evident that

- $\beta r = 1$ is still an important point.
- The electric field first reduces as β (or frequency) increases to the point $\beta r = 1$ and then changes to increase with β (or frequency) after this point.
- The magnetic field is a constant before the point $\beta r = 1$ and then increases with β (or frequency) at the same rate as the electric field when $\beta r > 1$.
- The ratio of **E**/**H** reduces (at the same rate as the electric field) first as β increases to about $\beta r = 1$; it then becomes a constant for $\beta r > 1$.

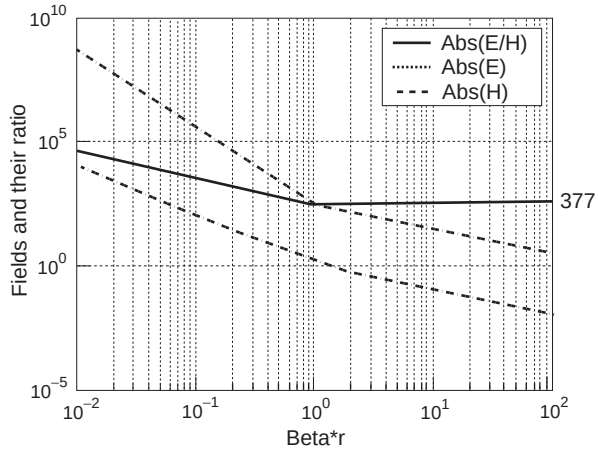


Figure 4.2 The fields **E**, **H** and **E/H** as a function of βr at a fixed frequency

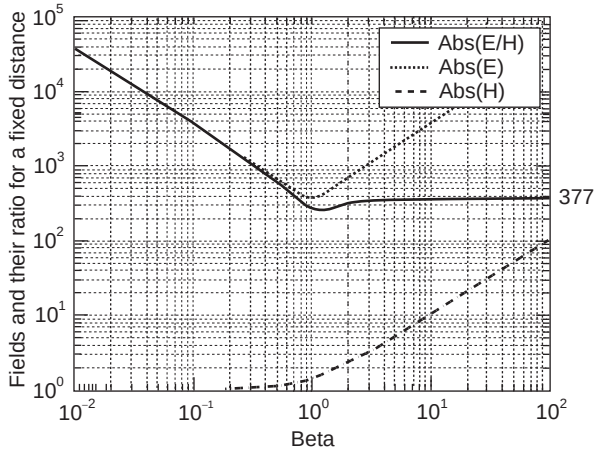


Figure 4.3 The fields **E**, **H** and **E/H** as a function of β at a fixed distance ($r = 1$)

These conclusions are very useful for applications such as radio frequency identification (RFID) and electromagnetic compatibility (EMC).

4.1.1.1 Far Field (Fraunhofer Region)

When $\beta r \gg 1$, i.e. $r \gg \frac{\lambda}{2\pi}$, we have $\frac{1}{\beta r} \gg \frac{1}{\beta^2 r^2} \gg \frac{1}{\beta^3 r^3}$. Equation (4.5) can therefore be simplified to

$$\begin{aligned} E_\theta &= \frac{jI\Delta l}{4\pi r}\eta\beta \sin\theta e^{-j\beta r} \\ E_r &\approx 0; \qquad E_\phi = 0 \end{aligned} \tag{4.6}$$

and the magnetic field is reduced to

$$\begin{aligned} H_\phi &= \frac{jI\Delta l}{4\pi r}\beta \sin\theta e^{-j\beta r} \\ H_r &= 0; \qquad H_\theta = 0 \end{aligned} \tag{4.7}$$

The fields are now simpler. It is important to note the following points:

- there is now just one electric field component and one magnetic field component;
- both fields are inversely proportional to the distance r;
- the electric field and magnetic field are orthogonal to each other and the cross product of these two is the power density function, which is inversely proportional to the distance squared r^2:

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}^* = \hat{\boldsymbol{r}} \left(\frac{I\Delta l}{4\pi r}\beta \sin\theta \right)^2 \eta \tag{4.8}$$

- the ratio of **E**/**H** is η, *the intrinsic impedance*, which is the same as that of the plane wave in Chapter 2;
- the fields are proportional to $\sin\theta$. They are zero at $\theta = 0^\circ$ and 180°, but maximum at $\theta = 90^\circ$.

Comparing the far field with the plane wave, we can see that they are basically the same except that the far-field amplitude is inversely proportional to the distance whilst the amplitude of the plane wave is constant. Thus, the far field can be considered a *local plane wave.*

It should be pointed out that the far-field condition is actually not that straightforward. The condition of $r \gg \lambda/2\pi$ was introduced for electrically small antennas and is just a function of the frequency and not linked to the antenna dimensions. When the antenna size D is electrically large, $D > \lambda$, the common definition of the *far-field condition* is

$$r > \frac{2D^2}{\lambda} \tag{4.9}$$

This is obtained under the condition that the maximum phase difference from any point on the antenna to the receiving point is less than $\pi/8$.

If an antenna cannot be considered electrically large, the recommended far-field condition is

$$r > 3\lambda \gg \lambda/2\pi \tag{4.10}$$

and some people use 10λ or other conditions which are also aimed at ensuring that the field is really far enough to be considered a local plane wave. For example, a half-wavelength dipole has maximum dimension $D = 0.5\lambda$, $2D^2/\lambda = 0.5\lambda$ is still too small to be considered far field. Thus, this additional condition is necessary.

The far-field condition is defined with certain ambiguity. Care must be taken when conducting antenna far-field measurements. If the separation between the transmitting and receiving antennas is not large enough, significant errors could be generated.

4.1.1.2 Near Field

When the far-field conditions are not met, the field is considered *near field*. The near field of a current element is illustrated in Figure 4.4. Since both E_θ and E_r have a number of terms linked to the frequency and distance and they are comparable, Equation (4.5) cannot be simplified. E_θ reaches the maximum when $\theta = 90^\circ$ while E_r peaks when $\theta = 0^\circ$ and 180° for a fixed distance.

The region for $r < \lambda/2\pi$ is normally called the *reactive near field*. The field changes rapidly with distance. From its power density function $\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}^*$, we can see that

- It contains both the radiating energy (the real part) and reactive energy (the imaginary part – it does not dissipate energy, which is like a capacitor or inductor). The latter is normally dominant in this region.
- It has components in the r and ϕ directions. The former is radiating away from the source and the latter is reactive.

Also, as shown in Figure 4.2, the electric field strength is much greater than the magnetic field strength (their ratio $> \eta$) when it is close to the source.

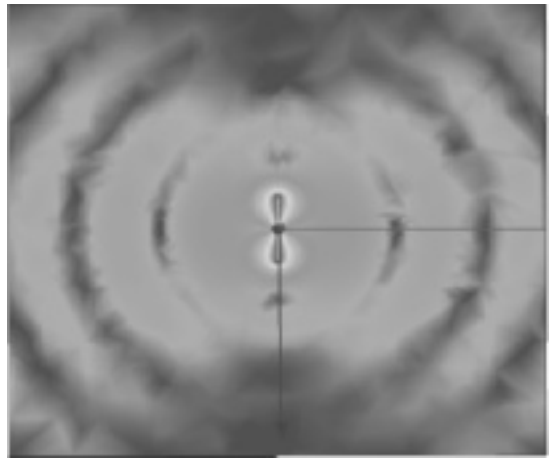


Figure 4.4 The electric field around a dipole antenna (computer simulation)

The region between the reactive near field and the far field is a transition region and is known as the *radiating near field* (or *Fresnel region*) where the reactive field becomes smaller than the radiating field. A brief summary of the three field regions is given in Table 4.1 and Figure 4.5. For EMC engineers, the near field is of great interest; but for antenna engineers, the far field is much more important than the near field.

4.1.2 Antenna Parameters from the Field Point of View

From the field point of view, we need to be able not only to distinguish the near field from the far field, but also characterize the field, especially the far field, so as to obtain more detailed information on the radiated field characteristics. The most important parameters include the radiation pattern, beamwidth, directivity, gain, efficiency factor, effective aperture, polarization and the bandwidth. The antenna temperature and radar cross-section are key parameters for some applications.

4.1.2.1 Radiation Pattern

The *radiation pattern* of an antenna is a plot of the radiated field/power as a function of the angle at a fixed distance, which should be large enough to be considered far field. The three-dimensional (3D) radiation pattern of the electrically short current element is plotted in Figure 4.6. The 3D pattern is an excellent illustration of the radiated field distribution as a function of angle θ and ϕ in space. Unfortunately, it is difficult and also very time-consuming

Table 4.1 Near-field and far-field conditions

Antenna size *D*	$D \ll \lambda$	$D \approx \lambda$	$D \gg \lambda$
Reactive near field	$r < \lambda/2\pi$	$r < \lambda/2\pi$	$r < \lambda/2\pi$
Radiating near field	$\lambda/2\pi < r < 3\lambda$	$\lambda/2\pi < r < 3\lambda$ and $2D^2/\lambda$	$\lambda/2\pi < r < 2D^2/\lambda$
Far field	$r > 3\lambda$	$r > 3\lambda$ and $2D^2/\lambda$	$r > 2D^2/\lambda$

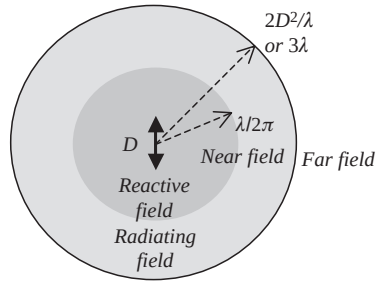


Figure 4.5 Radiated field regions of an antenna of maximum dimension D

to measure the 3D pattern of an antenna in practice. Most antennas have certain symmetrical features, thus, in reality, the most important patterns are the radiation patterns in the two main planes: the E-plane and the H-plane. The E-plane is the plane that the electric field $\boldsymbol{E}$ lies on, while the H-plane is the plane that the magnetic field $\boldsymbol{H}$ is on. For the ideal current element case, the electric field is E_θ and the magnetic field is H_ϕ, thus the E-plane pattern is the field E_θ measured as a function of θ when the angle ϕ and the distance are fixed, while the H-plane pattern is the field E_θ measured as a function of ϕ when the angle θ and the distance are fixed. The E-plane (at $\phi = 0$) and H-plane (at $\theta = \pi/2$) patterns of the short current element are shown in Figure 4.7. Obviously, this antenna has an *omnidirectional pattern* in the H-plane; this is a desirable feature for many mobile antennas since the antenna is not sensitive to orientation. Another special case is called the *isotropic antenna*, which has the same radiation power at all angles. This is a hypothetical case and cannot be realized in practice, but we do sometimes use it as a reference for analysis.

It should be pointed out that *the H-plane pattern is actually a measure of the electric field, not the magnetic field.* When we talk about the radiation pattern, it always means the electric field or the power (which is proportional to the electric field squared) pattern. When the patterns are plotted on a linear scale, the field pattern and power pattern may look very different. However, *when the patterns are plotted on a logarithmic scale (dB plot), both the normalized field and power patterns are the same* since $10\log(P/P_{\max})$ is the same as $20\log(E/E_{\max})$. Thus, in practice, we often plot the patterns in dB scale, which also makes it easy to see details of the field or power over a large dynamic range, especially some minor side lobes. It should be pointed out that if the electric field has more than one component, we normally need to plot each component pattern. The total radiation pattern is useful for gaining an overall view of the radiation.

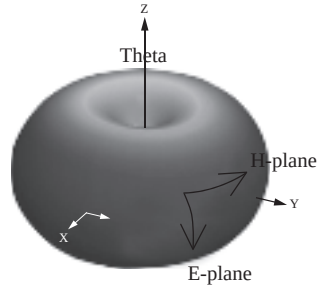


Figure 4.6 The 3D radiation pattern of an electrically short current element

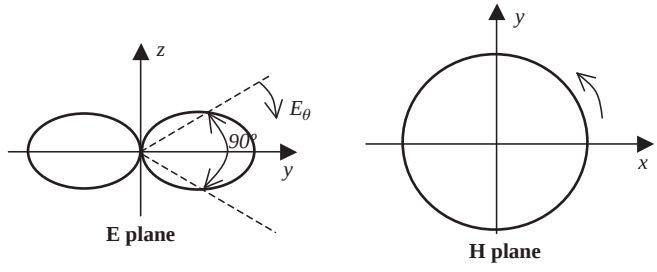


Figure 4.7 The E-plane and H-plane patterns of an electrically short current element

In addition to the 3D and 2D polar plots, another popular way to plot the radiation pattern is given in Figure 4.8 (also called the rectangular plot or universal plot). In this example, the horizontal axis is the rotation angle θ and the vertical axis is the radiated power in dB, which is normalized to the maximum – this is a very common practice since the radiation pattern is about the relative power distribution as a function of angle and the absolute value is not important as long as it is above the noise floor.

The radiation pattern contains a lot of useful information about the radiation characteristics of the antenna, some cannot be quantified (such as the shape of the pattern) and some can be quantified. Some of the most important ones are:

- the *half-power beamwidth* (HPBW) of the main lobe, also called the *3dB beamwidth* or just the *beamwidth* (to identify how sharp the beam is);
- the *10 dB beamwidth* or *first null beamwidth* (FNBW) (another one capturing the main beam shape);
- the *first side-lobe level* (expressed in dB, relative to the peak of the main beam);
- the *front-to-back ratio* (the peak of the main lobe over the peak of the back lobe, another attempt to identify the directivity of the antenna);
- null positions (sometimes used for anti-interference and positioning).

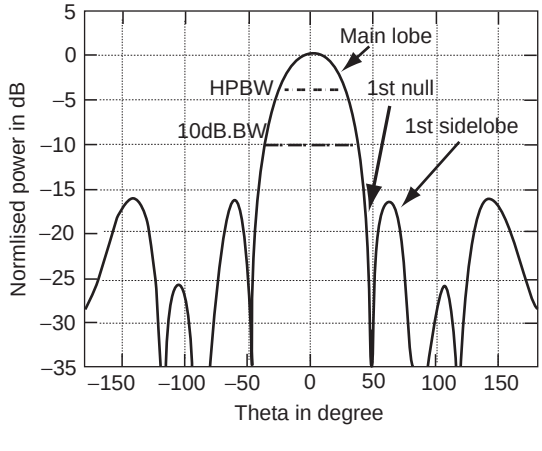


Figure 4.8 A radiation pattern illustrated in a conventional 2D plot

4.1.2.2 Directivity

From the field point of view, the most important quantitative information on the antenna is the *directivity*, which is a measure of the concentration of radiated power in a particular direction. It is defined as the ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions. The average radiation intensity is equal to the total radiated power divided by 4π. If the direction is not specified, the direction of maximum radiation is implied. Mathematically, the directivity (dimensionless) can be written as

$$D = \frac{U(\theta,\phi)}{U(\theta,\phi)_{av}} = \frac{4\pi U(\theta,\phi)}{P_t} = \frac{4\pi U(\theta,\phi)}{\oiint_{\Omega} U d\Omega} \tag{4.11}$$

where P_t is the total radiated power in W and U is the *radiation intensity* in W/unit solid angle. It is linked to the averaged radiated power density S_{av} (it has a unit of W/m^2) by distance squared, that is

$$U = r^2 S_{av} \tag{4.12}$$

Example 4.1: Directivity. The radiated power density of the electrically short current element is given earlier as (Equation 4.8):

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}^* = \hat{\boldsymbol{r}} \left(\frac{I\Delta l}{4\pi r} \beta \sin\theta \right)^2 \eta$$

Determine the directivity of the antenna as a function of the directional angles. Find the maximum directivity and express it in decibels.

Solution:
Since $S_{av} = S/2$ (time averaged), the radiation intensity is

$$U = r^2 \cdot \frac{1}{2} \left(\frac{I\Delta l}{4\pi r} \beta \sin\theta \right)^2 \eta = \frac{1}{2} \left(\frac{I\Delta l}{4\pi} \beta \right)^2 \eta \sin^2\theta = U_0 \sin^2\theta$$

where $U_0 = \frac{1}{2}(\frac{I\Delta l}{4\pi}\beta)^2\eta$. The total radiated power is given by

$$P_t = \oiint_{\Omega} U d\Omega = \int_0^{2\pi} \int_0^{\pi} U_0 \sin^2\theta \cdot \sin\theta d\theta d\phi = U_0 \cdot 2\pi \int_0^{\pi} \sin^3\theta d\theta$$

$$= U_0 \cdot 2\pi \cdot \frac{4}{3} = \frac{8\pi}{3} U_0$$

Thus, the directivity is

$$D = \frac{4\pi U}{P_t} = \frac{4\pi U_0 \sin^2\theta}{8\pi U_0/3} = 1.5 \sin^2\theta$$

The maximum directivity occurs at $\theta = \pi/2$, it is

$$D_0 = 1.5 = 1.76 \text{ dBi}$$

It is conventional to use the isotropic antenna as the reference, thus the directivity is normally expressed in *dBi*, and in this case the maximum directivity is 1.76 dBi. This means that the antenna radiates 1.76 dB more at the maximum direction than an isotropic antenna when the same amount of power is radiated. Sometimes, a half-wavelength dipole antenna is used as a reference; in this case, the directivity is expressed in *dBd*, where d is for dipole.

For a directional antenna with one main lobe, the directivity is closely linked to the half-power beamwidth in the E- and H- planes. If the beamwidths are known, there is an approximation which can be employed to calculate the directivity:

$$D \approx \frac{4\pi}{\theta_{HP}\phi_{HP}} \tag{4.13}$$

where θ_{HP} and ϕ_{HP} are the half-power beamwidths at the two principal orthogonal planes, expressed in radians. If they are given in degrees, this equation can be rewritten as

$$D \approx \frac{41253^\circ}{\theta_{HP}\phi_{HP}} \approx \frac{41000^\circ}{\theta_{HP}\phi_{HP}} \tag{4.14}$$

The validity of these two equations is based on the assumption that the pattern has only one major lobe. They are not suitable for calculating the directivity of, for example, the short current element, since this kind of antenna is not directional (but omnidirectional) in the H-plane and there is no HPBW in the H-plane.

4.1.2.3 Gain and Radiation Efficiency of Antennas

In practice, the total input power to an antenna can be obtained easily, but the total radiated power by an antenna is actually hard to get. The *gain* of an antenna is introduced to solve this problem. This is defined as the ratio of the radiation intensity in a given direction from the antenna to the total input power accepted by the antenna divided by 4π. If the direction is not specified, the direction of maximum radiation is implied. Mathematically, the gain (dimensionless) can be written as

$$G = \frac{4\pi U}{P_{in}} \tag{4.15}$$

where U is again the *radiation intensity* in W/unit solid angle and P_{in} is *the total input power accepted by the antenna* in W. It should be pointed out that the input power may be different from the input power accepted by the antenna when the feed line is not matched with the antenna. The voltage reflection coefficient Γ at the antenna input and the *match efficiency* $= 1 - |\Gamma|^2$ (Equation (4.41)) will be discussed later in the chapter. Basically, the feed line is matched with the antenna, $\Gamma = 0$ and the matching efficiency is 100%.

Comparing Equations (4.11) and (4.15), we can see that the gain is linked to the directivity by

$$G = \frac{P_t}{P_{in}} D = \eta_e D \tag{4.16}$$

Where η_e is called the *radiation efficiency factor of the antenna* and is the ratio of the radiated power to the input power accepted by the antenna:

$$\eta_e = \frac{P_t}{P_{in}} \tag{4.17}$$

This efficiency factor has taken both the conductor loss and dielectric loss into account, but not the impedance mismatch between the feed line and the antenna. Further discussion, from the circuit point of view, will be presented in the next section, which includes another formula to calculate this factor.

Example 4.2: Gain and radiated power. If the efficiency of the antenna in Example 4.1 is 50%, the VSWR at the antenna input is 3 and the input/supplied power is 1 W, find:

a. the power gain;
b. the total radiated power

Solution:

a. From Example 4.1, we know the directivity $D = 1.5$. Since the efficiency factor $\eta_e = 50\% = 0.5$, the power gain

$$G = \frac{P_t}{P_{in}} D = \eta_e D = 0.75 \text{ or } -1.24 \text{ dBi.}$$

The gain is smaller than 1 (or 0 dBi). This results from the low efficiency factor, which is very common for electrically small antennas.

b. If the VSWR = 3, we can easily obtain the voltage reflection coefficient using Equation (2.42) as

$$\Gamma = \frac{3-1}{3+1} = 0.5$$

this means a matching efficiency of $1 - 0.5^2 = 0.75 = 75\%$. Since the input power is 1 W, the total input power accepted by the antenna is 0.75 W and the total radiated power:

$$P_t = \eta_e P_{in} = 0.5 \times 0.75 = 0.375(W) = -4.26 \text{ dB}$$

4.1.2.4 EIRP

The antenna gain is often incorporated into a parameter called the *effective isotropic radiated power*, or *EIRP*, which is the amount of power that would have been radiated by an isotropic

antenna to produce the peak power density observed in the direction of maximum antenna gain, that is

$$EIRP = P_t G \tag{4.18}$$

where P_t is the radiated power. The EIRP is often stated in dBi. The advantage of expressing the power in terms of EIRP is that the pathloss between the transmitting antenna and the receiving antenna can be obtained easily by the ratio of the transmitted to the received EIRP; this is why in the mobile radio industry EIRP is widely used.

From the EIRP and the knowledge of the gain of an antenna, it is possible to calculate its radiated power and field strength values. For example, if the antenna's $EIRP = 10$ dBW (or just say 10 dB in practice) and the gain $G = 7$ dBi, the radiated power from the antenna is 3 dB.

Another closely related term is the *effective radiated power*, *ERP*, which is also widely used in the industry. The radiated power is calculated using a half-wavelength dipole rather than an isotropic antenna as reference, thus

$$ERP(dBW) = EIRP(dBW) - 2.15dBi \tag{4.19}$$

Example 4.3: EIRP and ERP. Find the EIRP and ERP of the antenna in Example 4.2

Solution:
Since the gain $G = 0.75$ and the radiated power $P_t = 0.375$, we have

$$EIRP = P_t G = 0.375 \times 0.75 = 0.2813 = -5.5 \text{ dB}$$
$$ERP(dBW) = EIRP(dBW) - 2.15dBi = -7.65 \text{ dB}$$

4.1.2.5 Effective Aperture and Aperture Efficiency

A common question about antennas is how the antenna directivity and gain are linked to its physical dimensions. Since the field/current on the antenna aperture is not uniform, the concept of antenna effective aperture is introduced to serve this purpose. The *effective aperture* A_e is less than the physical aperture A_p; the *aperture efficiency* is defined as the ratio of these two, that is

$$\eta_{ap} = \frac{A_e}{A_p} \tag{4.20}$$

This is normally in the range of 50 to 80%, and the effective aperture can be found as

$$A_e = \frac{\lambda^2}{4\pi} D \tag{4.21}$$

Thus, it is proportional to the directivity of the antenna, and the directivity can also be expressed in terms of the aperture size and aperture efficiency:

$$D = \frac{4\pi}{\lambda^2} A_e = \frac{4\pi}{\lambda^2} \eta_{ap} A_p \tag{4.22}$$

If the power density S at the receiving antenna is known, we can use the following equation to estimate the received power:

$$P_r = SA_e \tag{4.23}$$

Example 4.4: Effective aperture. The directivity of a pyramidal horn antenna of aperture width a and height b is

$$D = 6.4\frac{ab}{\lambda^2}$$

Find its aperture efficiency. If the power density around the antenna is 1 W/m^2, find the received power.

Solution:
Its effective aperture is

$$A_e = \frac{\lambda^2}{4\pi} D = \frac{6.4}{4\pi} ab = 0.5093ab$$

Its physical aperture $A_p = ab$, thus the aperture efficiency is

$$\eta_{ap} = \frac{A_e}{A_p} = 50.93\%$$

If the incoming power density is 1 W/m^2, the received power is

$$P_r = SA_e = 0.5093ab \text{ (W)}$$

This means the larger the aperture, the larger the received power.

4.1.2.6 Effective Height and the Antenna Factor

If an antenna is a wire-type antenna (such as a dipole), the *effective height* (h_e) may be used to replace the effective aperture. The effective height is proportional to the square root of the effective aperture. It is defined as the ratio of induced open-circuit voltage V on the antenna to the incident electric field E or

$$h_e = \frac{V}{E} \propto \sqrt{A_e}\,(\text{m}) \tag{4.24}$$

Another closely related parameter is the *antenna factor* (A_F), which is defined as the ratio of the incident electric field E to the induced voltage V_0 at the antenna terminal when it is connected to a load/cable (50 ohms by default). This induced voltage is different from the V in Equation (4.27). Thus

$$A_F = \frac{E}{V_0}(1/\text{m}) \tag{4.25}$$

This links the radiated/received field strength to the voltage at the antenna terminal and is a very useful parameter for field strength and voltage conversion in field measurement applications. Thus, it is a widely used device descriptor in the EMC area. However, it is not part of standard antenna terminology. The antenna factor reflects the use of an antenna as a field measuring device or probe. It converts the reading of voltage on equipment (such as a receiver) to the field at the antenna. It is also linked to the gain and effective aperture of the antenna. More information on the antenna factor can be found in [3, 4].

4.1.2.7 Antenna Temperature

The radiation from different sources is intercepted by antennas and appears at their terminals as an antenna temperature, which is defined as [5]

$$T_A = \frac{\int_0^{2\pi}\int_0^{\pi} T_B(\theta,\phi)G(\theta,\phi)\cdot\sin\theta d\theta d\phi}{\int_0^{2\pi}\int_0^{\pi} G(\theta,\phi)\cdot\sin\theta d\theta d\phi} \tag{4.26}$$

where T_B is the source brightness temperature of the radiation source and G is the antenna gain pattern.

When an antenna is used for receiving purposes, the antenna temperature may be required for system considerations and signal-to-noise ratio (SNR) calculation. In this case, the source is remote (such as the sky or Mars) and the antenna is just a load to the system and may be viewed as a remote temperature-measuring device. But this temperature has nothing to do with the physical temperature of the antenna if the antenna is a lossless device; the temperature is equal to the source (such as the distant sky) temperature. In reality, the antenna has certain loss. An extreme case would be one where the antenna is completely lossy, here the antenna temperature is equal to its physical temperature. But, generally speaking, its physical temperature is partially linked to the antenna temperature. The larger the effective aperture/gain, the larger the antenna temperature.

4.1.2.8 Friis's Transmission Formula and the Radar Cross-Section

Another use of the aperture concept is for the derivation of the Friis transmission formula, which relates the power received to the power transmitted between two antennas separated by a distance r, which is large enough to be considered in the antennas' far field. As shown in Figure 4.9, from Equations (4.11) and (4.12), the radiated power density from an antenna with

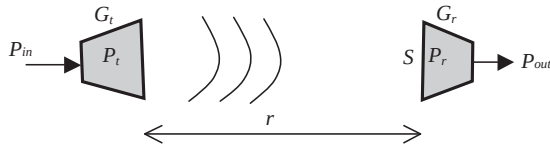


Figure 4.9 Transmitting and receiving antennas for the Friis transmission formula

gain G_t can be expressed as

$$S = \frac{P_{in} G_t}{4\pi r^2} \tag{4.27}$$

From Equations (4.21) and (4.23), we can obtain the received power

$$P_r = \frac{P_{in} G_t}{4\pi r^2} \cdot \frac{\lambda^2}{4\pi} D_r \tag{4.28}$$

To obtain the output power from the receiving antenna, we must take the receiving antenna's efficiency factor η_{er} into account, thus

$$P_{out} = P_r \eta_{er} = \frac{P_{in} G_t}{4\pi r^2} \cdot \frac{\lambda^2}{4\pi} D_r \eta_{er} = \frac{P_{in} G_t}{4\pi r^2} \cdot \frac{\lambda^2}{4\pi} G_r = \left(\frac{\lambda}{4\pi r}\right)^2 G_t G_r \cdot P_{in} \tag{4.29}$$

This is the *Friis transmission formula* and it links the input power of the transmitting antenna to the output power of the receiving antenna. $(\lambda/4\pi r)^2$ is called the *free space loss factor* – resulting from the spherical spreading of the radiated energy. This formula can also be expressed using the effective aperture of both antennas (A_{et} and A_{er}) as

$$P_{out} = \eta_{et} \eta_{er} \frac{A_{et} A_{er}}{\lambda^2 r^2} \cdot P_{in} \tag{4.30}$$

This equation has been widely used for communications and radar applications.

For radar applications, Friis's equation can be related to the *radar cross-section (RCS)*, σ, which is defined as the ability of a target to reflect the energy back to the radar and is the ratio of the backscattered power to the incident power density; that is

$$\sigma = \frac{\text{scattered power}}{\text{incident power density}} \tag{4.31}$$

It is a far-field parameter and is expressed in m^2 or dBsm (decibels per square meter). The typical RCS for a jumbo jet aircraft is about 100 m^2, for a fighter aircraft about 4 m^2, for an adult man 1 m^2 and for a bird 0.01 m^2. It is not a function of frequency.

The power density in Equation (4.27) can be considered the incident power density on the target, and the part reflected back to the radar is $\sigma S/4\pi r^2$ and the output power at the

radar antenna is

$$P_{out} = \eta_e A_e \cdot \sigma S/4\pi r^2 = \frac{\sigma \lambda^2}{(4\pi)^3 r^4} G_t^2 \cdot P_{in} \tag{4.32}$$

Thus, the received power is proportional to λ^2, but now inversely proportional to r^4, not r^2.

4.1.2.9 Polarization

The antenna polarization is the same as the polarization of its radiating wave. As discussed in Chapter 3, there are basically three types of polarization: linear, circular and elliptical. Which polarization is generated depends on how the current moves in the antenna. For a linear polarization, the current should travel along one axis; for a circular polarization, two orthogonal currents with 90-degree phase offset should be created on the antenna. In practice, mixed polarizations may be found in many antennas since an antenna has to meet many requirements. Trade-offs may have to be made and as a result a pure linearly polarized or circularly polarized antenna may not be possible or necessary. For example, most of the mobile phone antennas are not purely linearly polarized since they are employed to receive signals with mixed polarizations. However, for a line-of-sight communication system, the polarization has to be matched in order to achieve the maximum efficiency of the whole system. Two orthogonally polarized antennas cannot communicate with each other due to polarization mismatch. More discussion will be provided in Section 5.4.3.

4.1.2.10 Bandwidth

Many antenna parameters are functions of frequency. When the frequency is changed, the radiation pattern may also be changed, which may result in changes to the directivity, gain, HPBW and other parameters. Thus, it is important to ensure that the right parameters are chosen when the antenna bandwidth is considered.

4.2 Antennas to Transmission Lines

4.2.1 Antenna Parameters from the Circuit Point of View

As we mentioned earlier, the antenna is a transition device linking the transmission line and radio waves; thus, from the circuit point of view, the antenna is just a load to a transmission line. As shown in Figure 4.10, the most important parameter is, therefore, its impedance.

4.2.1.1 Input Impedance and Radiation Resistance

Antenna input impedance (Z_a) is defined as the impedance presented by an antenna at its terminals or the ratio of the voltage to current at its terminals. Mathematically, the input impedance is

$$Z_a = \frac{V_{in}}{I_{in}} = R_a + jX_a \tag{4.33}$$

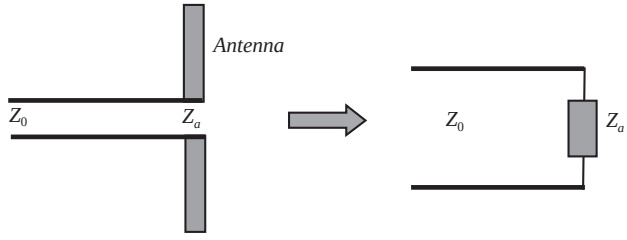


Figure 4.10 Antenna input impedance model

where V_{in} and I_{in} are the input voltage and current at the antenna input, respectively. The input impedance is a complex number; the real part consists of two components:

$$R_a = R_r + R_L \tag{4.34}$$

where R_r is the radiation resistance and R_L is the loss resistance of the antenna.

The *radiation resistance* is equivalent to the resistance which would dissipate the same amount of power as the antenna radiates when the current equals the input current at the terminals. If the total radiated power is P_t, then

$$R_r = \frac{2P_t}{I_{in}^2} \tag{4.35}$$

The loss resistance is from conductor loss and dielectric loss if lossy materials are used.

If the antenna is directly connected to a source of impedance Z_s, for a matched load the following condition should be met:

$$Z_s = Z_a^* = R_a - jX_a \tag{4.36}$$

In reality, the antenna is normally connected to a short transmission line with a standard characteristic impedance of 50 Ω or 75 Ω. Thus, the desired antenna input impedance is 50 Ω or 75 Ω.

4.2.1.2 Reflection Coefficient, Return Loss and VSWR

Impedance matching is extremely important. Since the antenna is just a load to a transmission line from the circuit point of view, we know that the reflection coefficient, return loss and voltage standing wave ratio (VSWR) can be used to judge how well a load is matched with the transmission line, from the transmission line theory in Chapter 2. All these three parameters are interlinked. The specific representations are as follows:

- reflection coefficient (a complex number):

$$\Gamma = \frac{Z_a - Z_0}{Z_a + Z_0} \tag{4.37}$$

- return loss (expressed in dB):

$$L_{RT} = -20\log_{10}(|\Gamma|) = -20\log_{10}\left|\frac{Z_a - Z_0}{Z_a + Z_0}\right| \tag{4.38}$$

- VSWR (a ratio between 1 and infinity)

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{4.39}$$

The commonly required specification of an antenna is $L_{RT} > 10$ dB or $VSWR < 2$. These two are actually very close, as seen in Table 2.1. For mobile phone antennas, it is also common to specify $VSWR < 3$ or $L_{RT} > 6$ dB.

4.2.1.3 Radiation Efficiency, Matching Efficiency and Total Efficiency

The radiation efficiency was defined by Equation (4.17) and can be calculated using the radiated power and the input power accepted by the antenna. Since they are linked directly to the radiation resistance and loss resistance, the radiation efficiency factor can also be calculated using the following equation

$$\eta_e = \frac{P_t}{P_{in}} = \frac{R_r}{R_r + R_L} \tag{4.40}$$

From the efficiency point of view, the larger the radiation resistance, the larger the efficiency factor. *But from the system matching point of view, we need the radiation resistance to match with the characteristic impedance of the line and the loss resistance to be zero if possible*. This is the fundamental requirement for an antenna from the circuit point of view.

If the input impedance of an antenna is not matched with the feeder impedance, then a part of the signal from the source will be reflected back; there is a reflection (mismatch) loss. This is characterized by the *matching efficiency* (also known as *reflection efficiency*), as mentioned earlier, which is defined as the ratio of the input power accepted by the antenna to the source supplied power (P_s):

$$\eta_m = \frac{P_{in}}{P_s} = 1 - |\Gamma|^2 \tag{4.41}$$

Thus, the total efficiency of the antenna system (feed and the antenna) is the product of the two efficiencies:

$$\eta_t = \frac{P_t}{P_s} = \eta_m \eta_e \tag{4.42}$$

This has taken the feed-antenna mismatch into account.

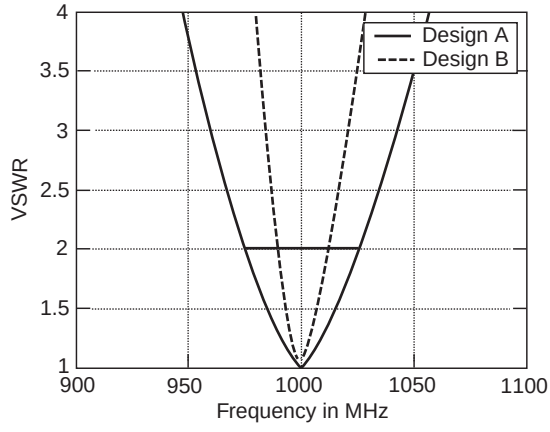


Figure 4.11 VSWR plotted against the frequency for two designs

4.2.1.4 Bandwidth

From the circuit point of view, the antenna bandwidth is a simple but important parameter. It is normally defined as the frequency bandwidth with $L_{RT} > 10$ dB or $VSWR < 2$. Take Figure 4.11 as an example, which is a zoomed version of Figure 2.19; the center frequency of the antenna is 1000 MHz. For the required $VSWR < 2$, the bandwidth for Design A is from 975 MHz to 1026 MHz, whilst the bandwidth for Design B is from 990 MHz to 1012 MHz, and the relative bandwidths for these two designs are 5.1% and 2.2%, respectively. Both are narrowband.

The antenna bandwidth is a very special parameter since we need to consider the parameter from both the circuit point of view (*VSWR* or L_{RT}) and the field point of view (radiation pattern). It is normally easy to meet one of the requirements but difficult to meet both requirements – this is particularly true for a broadband antenna, which normally means that its VSWR and radiation pattern have a relative bandwidth of at least 20%. In some applications, a much wider bandwidth is required.

4.3 Summary

In this chapter we have introduced the basic theory of antennas – the essence of antenna design is how to control the current distribution, which determines the radiation pattern and input impedance. A detailed discussion on antenna near fields and far fields has been conducted. All the important parameters of an antenna, which have been grouped from the field point of view and the circuit point of view, have been defined and discussed. A summary of these parameters is given in Figure 4.12. The Friis transmission formula and RCS have also been addressed.

It is worth pointing out that an antenna can be viewed as a two-port network, thus the field quantities (such as the radiation pattern) are related to the transmission coefficient (S_{21}) while the circuit quantities (such as the input impedance and VSWR) are determined by the reflection coefficient (S_{11}) of the network.

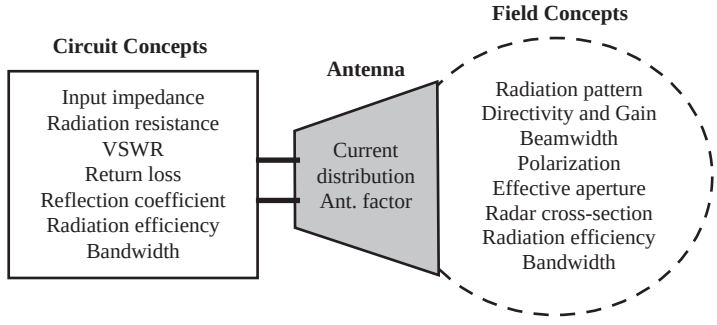


Figure 4.12 A summary of the most important antenna parameters

References

[1] R. E. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, Inc., 1985.
[2] J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd edition, McGraw-Hill, Inc., 2002.
[3] A. A. Smith, Jr., 'Standard-site method for determining antenna factors,' *IEEE Trans. Electromagnet. Compat.*, **24**, 316–322, 1982.
[4] R. P. Clayton, *Introduction to Electromagnetic Compatibility*, Wiley Interscience, 1992, pp. 202–206.
[5] C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd edition, John Wiley & Sons, Inc., 1997.

Problems

Q4.1 If you were asked to use one equation to explain the antenna theory, which one would it be? Why?
Q4.2 Explain how a radio wave may be generated.
Q4.3 What are the most important antenna parameters?
Q4.4 An ideal element current source is given as $\boldsymbol{J} = \hat{\boldsymbol{z}} I \Delta l$. Derive its radiated electric field using spherical coordinates.
Q4.5 What are the near field and far field of an antenna? What are their major differences?
Q4.6 Find the far-field condition of a half-wavelength dipole.
Q4.7 Find the far-field condition of a satellite dish antenna of dimension 60 cm operating at 12 GHz.
Q4.8 Explain the concept of the *antenna factor*. If the electric field around an antenna is 1 V/m and its antenna factor is 2/m, find the voltage at its matched output.
Q4.9 The radiated power density of an antenna in its far field can be expressed as:

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}^* = \hat{\boldsymbol{r}} \left(\frac{C}{r} \sin\theta \sin\phi \right)^2, \text{ for } 1 < \phi < \pi$$

where C is a coefficient determined by the frequency and antenna dimensions. Find the directivity of the antenna as a function of the directional angles. Find the maximum directivity and express it in decibels.

Q4.10 A monostatic radar is employed to detect a target of a typical RCS of 4 m^2. Its operational frequency is 10 GHz, the transmitted power is 1000 W. If its antenna gain is 30 dBi, find the operational range/distance of this radar.

Q4.11 Compare the usefulness of the reflection coefficient, VSWR and return loss for antenna characterization.

Q4.12 An antenna has an input impedance of $73 + j10$ ohms. If it is directly connected to a 50 ohm transmission line, find the reflection coefficient, VSWR and return loss.

5

Popular Antennas

Chapter 4 has covered the basics of antennas and laid down the foundation for us to gain a better understanding of antennas. In this chapter we are going to examine and analyze some of the most popular antennas using relevant theories, to see why they have become popular, what their major features and properties (including advantages and disadvantages) are and how they should be designed.

Since the start of radio communications over 100 years ago, thousands of antennas have been developed. They can be categorized by various criteria:

- in terms of the bandwidth, antennas can be divided into narrowband and broadband antennas;
- in terms of the polarization, they can be classified as linearly polarized or circularly polarized antennas (or even elliptically polarized antennas);
- in terms of the resonance, they can be grouped as resonant or traveling wave antennas;
- in terms of the number of elements, they can be organiz as element antennas or antenna arrays.

In this book we separate them according to their physical structures into wire-type antennas and aperture-type antennas. This is because different types of antenna exhibit different features and can be analyzed using different methods and techniques. Since antenna arrays can be formed by both types and possess some special features, they will be discussed in Section 5.3.

5.1 Wire-Type Antennas

Wire-type antennas are made of conducting wires and are generally easy to construct, thus the cost is normally low. Examples include dipoles, monopoles, loops, helices, Yagi–Uda and log-periodic antennas

5.1.1 Dipoles

Dipoles are one of the simplest but most widely used types of antenna. Hertz used them for his famous experiment. As shown in Figure 5.1, a dipole can be considered a structure evolved

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

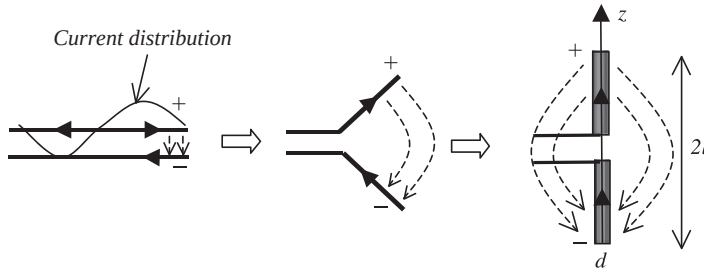


Figure 5.1 Evolution of a dipole of total length $2l$ and diameter d

from an open-end, two-wire transmission line. A typical structure of a dipole consists of two metal wires which are normally of equal length.

In Chapter 4, an ideal short element of constant current was discussed. In reality, we cannot find such an antenna since the current is zero at the end of a line. As we now understand, for antenna analysis and design, it is all about the current distribution. Once the current distribution is known, other parameters, such as the radiation pattern and input impedance, can be obtained.

5.1.1.1 Current Distribution

For an open-end transmission line, the reflection coefficient at the end ($z = 0$) is 1. From Equation (2.20), the current distribution along such a line can be expressed as

$$I(z,t) = \mathrm{Re}\left(\frac{1}{Z_0}(A_1 e^{j(\omega t-\beta z)} - A_1 e^{j(\omega t+\beta z)})\right) = \frac{2}{Z_0} A_1 \sin(\omega t)\sin(\beta z) \tag{5.1}$$

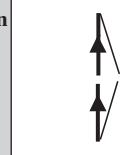
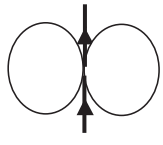
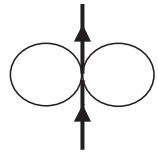
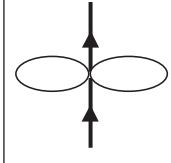
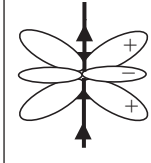
This is a standing wave with peaks at $z = \lambda/4 + n\lambda/2$ (n are integers). If the dipole is very thin (ideally zero diameter), the current distribution can be approximated by this equation. That is, if the coordinate origin is chosen at the center of the dipole in Figure 5.1, the current can be expressed as

$$I(z) = \begin{cases} I_0 \sin(\beta(l-z)), & 0 \le z \le l \\ I_0 \sin(\beta(l+z)), & -l \le z \le 0 \end{cases} \tag{5.2}$$

where I_0 is the *maximum possible current* (different from the current at the feeding point) on the line. If the dipole length is shorter than $\lambda/2$, the maximum current on the dipole is less than I_0. This could be very confusing. The time term $\sin(\omega t)$ is omitted from the representation. Thus, the current along the dipole is a time-varying standing wave. From Equation (5.2), we can see that the current is

- zero at the end of the pole;
- $I_0 \sin(\beta l)$ at the feeding point;
- a periodic function with a period equal to one wavelength.

Table 5.1 Summary of some dipole characteristics

Dipole length 2*l*	$\lambda/10$	$\lambda/2$	λ	1.5λ
Current distribution				
Radiation pattern				
Directivity	**1.5 or 1.76 dBi**	**1.64 or 2.15 dBi**	**2.4 or 3.8 dBi**	**About 2.3**
HPBW	**90°**	**78°**	**47°**	**NA**
Input impedance	***R*: very small (~2 Ω)** ***jX*: capacitive**	***R*: ~73 Ω** ***jX*: ~ 0 Ω**	***R*: very large** ***jX*: ~ 0 Ω** **for thin dipole**	***R*: ~ 100 Ω** ***jX*: ~ 0 Ω** **for thin dipole**
Note	***jX* sensitive to the radius**	***R*+*jX* not sensitive to the radius**	***R*+*jX* sensitive to the radius**	***R*+*jX* sensitive to the radius**

The current distributions for the dipole of length $2l = \lambda/10, \lambda/2, \lambda$ and 1.5λ are shown in Table 5.1. Note that the current distributions on both poles are symmetrical (even function as shown in Equation (5.2)) about the antenna center.

5.1.1.2 Radiation Pattern

Replacing the current density in Equation (4.3) by Equation (5.2) gives

$$\boldsymbol{E}(\boldsymbol{r}) = \left(-j\omega\mu + \frac{\nabla\nabla\bullet}{j\omega\varepsilon}\right)\int_{-l}^{l} \hat{\boldsymbol{z}} I(z)\frac{e^{-j\beta r + j\beta z\cos\theta}}{4\pi r}dz \tag{5.3}$$

After some mathematical manipulation, the radiated electric field in the far field can be obtained approximately as

$$E_\theta \approx j\eta\frac{I_0 e^{j\beta r}}{2\pi r}\left(\frac{\cos(\beta l\cos\theta) - \cos(\beta l)}{\sin\theta}\right) \tag{5.4}$$

and is plotted in Table 5.1. Here η is the intrinsic impedance.

The magnetic field can be obtained as

$$H_\phi = \frac{E_\theta}{\eta} \approx j\frac{I_0 e^{j\beta r}}{2\pi r}\left(\frac{\cos(\beta l\cos\theta)-\cos(\beta l)}{\sin\theta}\right) \tag{5.5}$$

Thus, the averaged power density is

$$\boldsymbol{S}_{\text{av}} = \frac{1}{2}Re\left(\boldsymbol{E}\times\boldsymbol{H}^*\right) = \hat{\boldsymbol{r}}\frac{\eta I_0^2}{8\pi^2 r^2}\left(\frac{\cos(\beta l\cos\theta)-\cos(\beta l)}{\sin\theta}\right)^2 \tag{5.6}$$

Using Equation (4.12), the radiation intensity is

$$U = r^2 S_{\text{av}} = \frac{\eta I_0^2}{8\pi^2}\left(\frac{\cos(\beta l\cos\theta)-\cos(\beta l)}{\sin\theta}\right)^2 \tag{5.7}$$

When βl is small ($< \pi/4$), $\cos(\beta l) \approx 1-(\beta l)^2/2$ and we have

$$\left(\frac{\cos(\beta l\cos\theta)-\cos(\beta l)}{\sin\theta}\right) \approx \frac{1}{2}(\beta l)^2\sin\theta$$

Thus, for short dipoles, the radiated field and radiation intensity are

$$\begin{aligned} E_\theta &\approx j\eta\frac{I_0 e^{j\beta r}}{4\pi r}(\beta l)^2\sin\theta \approx j\eta\frac{I_{in}e^{j\beta r}}{4\pi r}(\beta l)\sin\theta \\ U &\approx \frac{\eta I_{in}^2}{32\pi^2}(\beta l)^2\sin^2\theta \end{aligned} \tag{5.8}$$

where $I_{in} = I_0\sin\beta l$ is the dipole input current and more useful than I_0 for short dipoles.

The E-plane radiation patterns for the length $2l = \lambda/10$, $\lambda/2$, λ and 1.5λ are shown in Table 5.1. It is apparent that the current distribution indeed determines the radiation pattern. When the dipole length is less than a wavelength, the currents on both poles have the same polarity and there is only one lobe on both sides and no side lobes. But when the length is greater than a wavelength, the currents on the dipole become complicated; they travel in two opposite directions, which results in a split of the radiation pattern. '+' and '−' in the plots indicate opposite directions of the current phase.

5.1.1.3 Directivity and Gain

Using Equation (4.11), we can calculate the directivity by

$$D = \frac{4\pi U(\theta,\phi)}{P_t} = \frac{4\pi U(\theta,\phi)}{\int_0^{2\pi}\int_0^{\pi} U\sin\theta d\theta d\phi} = \frac{2U(\theta,\phi)}{\int_0^{\pi} U\sin\theta d\theta} \tag{5.9}$$

It is not possible to find a simple expression for the directivity. A numerical solution can be found once the frequency and dipole length are known.

The gain of the antenna can be obtained by Equations (4.16) and (4.34). We need to know what material is used to make the antenna in order to calculate the gain. An example is given in Example 5.1.

The directivities for the dipole of length $2l = \lambda/10, \lambda/2, \lambda$ and 1.5λ are also shown in Table 5.1. The half-power beamwidths (HPBWs) for these antennas are given in the table as well. The directivity increases with the dipole length and reaches a first maximum (about 3.2) when the length is around 1.25λ (not shown in the table) [1, 2]. The relation between the directivity and beamwidth in Equation (4.13) cannot be applied to this kind of antenna. For some antennas, their HPBWs may be about the same, but their radiation patterns could actually be very different. Thus, it is essential to obtain the 2D or even 3D radiation pattern and not to rely purely on the HPBW to judge the radiation characteristics of an antenna.

5.1.1.4 Radiation Resistance and Input Impedance

Once the current distribution is obtained, we can calculate the radiation resistance using Equation (4.35); that is

$$R_r = \frac{P_t}{I_{in}^2} = \frac{2\pi \int_0^\pi U \sin\theta d\theta}{(I_0 \sin \beta l)^2} \tag{5.10}$$

We can find a complicated but explicit expression of the radiation resistance as

$$\begin{aligned} R_r = {} & \frac{\eta}{2\pi}\{C_E + \ln(\beta l/2) - C_i(\beta l/2) + 0.5\sin(\beta l/2)[S_i(\beta l) - 2S_i(\beta l/2)] \\ & + 0.5\cos(\beta l/2)[C + \ln(\beta l/4) + C_i(\beta l) - 2C_i(\beta l/2)]\} \end{aligned} \tag{5.11}$$

where

$C_E = 0.5772$ is Euler's constant;

$C_i(x) = \int_\infty^x \frac{\cos y}{y} dy$ is the cosine integral;

$S_i(x) = \int_0^x \frac{\sin y}{y} dy$ is the sine integral.

This complex expression is not really suitable for applications. A lot of effort has been made to make it simpler and easier to use. One such attempt can be found in [3]. If the loss resistance is neglected, the input impedance of a dipole less than a wavelength can be approximated as

$$Z_a \approx f_1(\beta l) - j\left(120\left(\ln\frac{2l}{d} - 1\right)\cot(\beta l) - f_2(\beta l)\right) \tag{5.12}$$

where

$$\begin{aligned} f_1(\beta l) &= -0.4787 + 7.3246\beta l + 0.3963(\beta l)^2 + 15.6131(\beta l)^3 \\ f_2(\beta l) &= -0.4456 + 17.0082\beta l - 8.6793(\beta l)^2 + 9.6031(\beta l)^3 \end{aligned}$$

d is the diameter of the dipole.

This formula is valid only when the dipole length is not much longer than half a wavelength with an error up to 0.5 Ω. In practice, this is the most useful range. If the impedance for a longer dipole is to be calculated, the best approach is to use numerical computation (software packages are widely available on the market), which will be addressed in Chapter 6.

For an electrically short dipole, Equation (5.11) can be further simplified to

$$Z_a \approx 20(\beta l)^2 - j120\left(\ln\frac{2l}{d} - 1\right)/(\beta l) \tag{5.13}$$

This expression clearly indicates that

- The radiation resistance (the real part of Z_a in this case) is proportional to the square of the dipole length and inversely proportional to the wavelength squared.
- The reactance of the input impedance is a function of the radius and length of the dipole. The smaller the radius, the larger the amplitude of the reactance. It is proportional to the wavelength.

The input impedances for the dipole of length $2l = \lambda/10, \lambda/2, \lambda$ and 1.5λ are also shown in Table 5.1. It should be pointed out that the reactance of the input impedance is very sensitive to the radius of the dipole. The only exception is when the length is near half of the wavelength, the very first resonant length. Although $2l = \lambda$ and 1.5λ may also be considered resonant cases, they are very sensitive to a change in the radius. If the radius is not negligible, they are actually not resonant. A good illustration of the input impedance as a function of the normalized dipole length (and frequency) is given in Figure 5.2 [1], where three dipole antennas of different radii

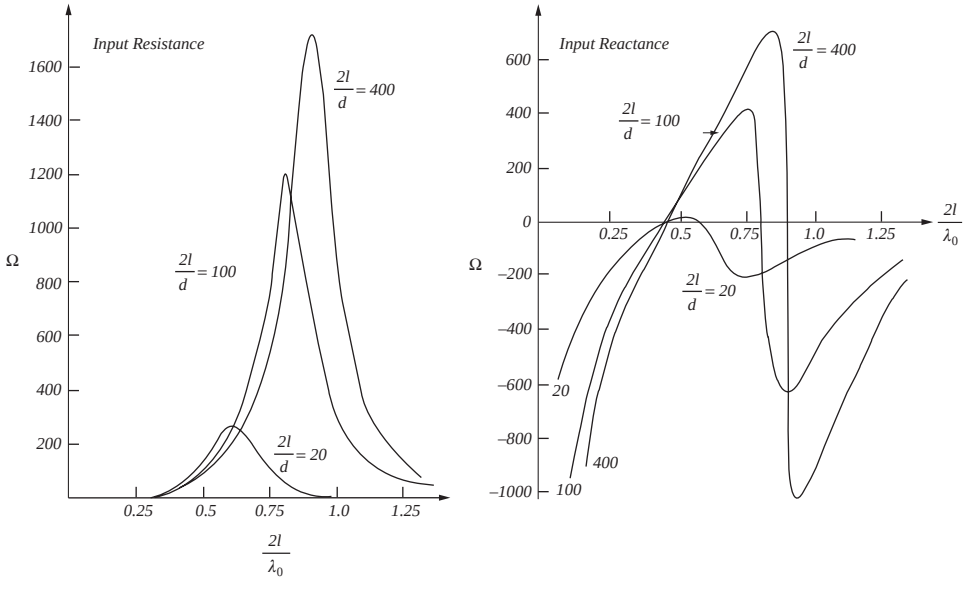


Figure 5.2 Input impedance as a function of the normalized dipole length. (R. E. Collin, Antennas and Radiowave Propagation, McGraw-Hill, Inc., 1985. Reproduced by permission of The McGraw-Hill Companies)

are given. We can see that

- All dipoles are resonant near half-wavelength (due to the non-zero radius, the resonant length is slightly shorter than $\lambda/2$), but not at one wavelength.
- The fatter (larger radius) the dipole, the broader the bandwidth.

5.1.1.5 Why the Half-Wavelength Dipole is the Most Popular Dipole

Based on the discussions above, we can conclude that the following four reasons make the $\lambda/2$ dipole very popular:

- Its radiation pattern is omnidirectional in the H-plane, which is required by many applications (including mobile communications).
- Its directivity (2.15 dBi) is reasonable – larger than short dipoles although smaller than that of the full-wavelength dipole.
- The antenna is longer than a short dipole but much shorter than the full-wavelength dipole, hence it is a good trade-off between the directivity and size.
- The input impedance is not sensitive to the radius and is about 73 Ω, which is well matched with a standard transmission line of characteristic impedance 75 Ω or 50 Ω (with $VSWR < 2$). This is probably the most important and unique reason.

Thus, for overall performance, the half-wavelength dipole antenna is a definite winner.

Example 5.1: Short dipole. A dipole of length $2l = 3$ cm and diameter $d = 2$ mm is made of copper wire ($\sigma = 5.7 \times 10^7$ S/m) for mobile communications. If the operational frequency is 1 GHz,

a. obtain its radiation pattern and directivity;
b. calculate its input impedance, radiation resistance and radiation efficiency;
c. if this antenna is also used as a field probe at 100 MHz for EMC applications, find its radiation efficiency again, and express it in dB.

Solution:
Since the frequency is 1 GHz, the wavelength $\lambda = 30$ cm, $2l/\lambda = 0.1$ and $\beta l = 0.1\pi$.

a. It is a short dipole, therefore we use Equation (5.8) to yield the radiation pattern. The 3D and the normalized E-plane patterns are shown in Figures 4.6 and 5.3 respectively. The H-plane pattern (ϕ) is omnidirectional. The HPBW in the E-plane is 90 degrees (from 45 to 135 degrees in the figure), as indicated in Table 5.1. This plot is almost exactly the same as that obtained by using Equation (5.7) with an error < 0.03 dB.
   Using Equations (5.8) and (4.9) we find directivity to be 1.5 or 1.76 dBi.

b. Equation (5.13) can be employed to obtain the input impedance, the result is

$$Z_a = 1.93 - j652\ \Omega$$

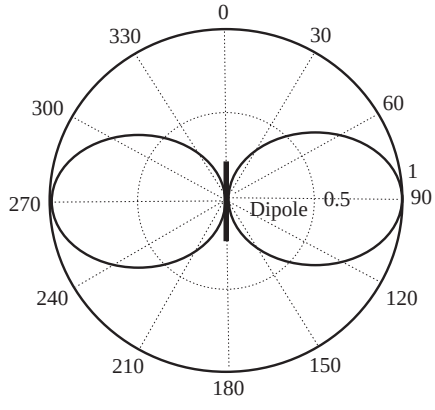


Figure 5.3 E_θ radiation pattern of a short dipole as a function of θ (E-plane)

(If (5.12) was used, the result would be $Z_a = 2.34 - j634\Omega$, which is slightly different).

To calculate the loss resistance, we first need to find out the skin depth. Using Equation (3.49), the skin depth in this case is

$$\delta \approx \sqrt{\frac{1}{\pi f \mu \sigma}} = \sqrt{\frac{1}{\pi \times 10^9 \times 4\pi \times 10^{-7} \times 5.7 \times 10^7}} = 2.1 \times 10^{-6}\,(\text{m})$$

It is much smaller than the radius of the wire. Thus, we need to use Equation (3.50) to compute the loss resistance of the dipole:

$$R_L \approx \frac{2l\sqrt{f\mu}}{d\sqrt{\pi\sigma}} \approx 0.04\,\Omega$$

Since Equation (5.13) has not taken the loss resistance into account, the more accurate input impedance is

$$Z_a = 1.97 - j652\,\Omega$$

and the radiation efficiency is

$$\eta_e = \frac{R_r}{R_r + R_L} = \frac{1.93}{1.93 + 0.04} = 97.97\%$$

c. The frequency is now 100 MHz and the wavelength $\lambda = 300$ cm, thus $2l/\lambda = 0.01$ and $\beta l = 0.01\pi$. This is an electrically small antenna; we can use the same approach as in b. to obtain:

The antenna input impedance (without ohmic loss): $Z_a = 0.0197 - j6524.3\ (\Omega)$
The skin depth: $\delta \approx 6.67 \times 10^{-6} < d/2$
The loss resistance: $R_L \approx 0.0126\ (\Omega)$

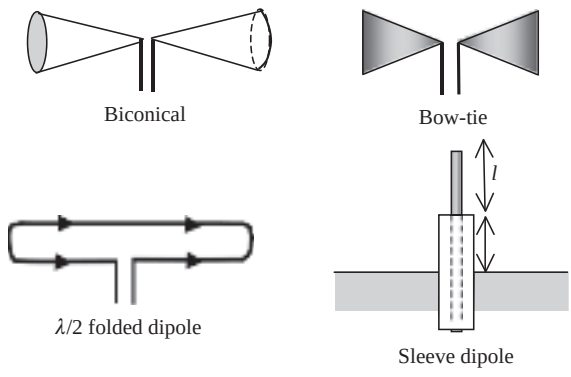


Figure 5.4 Some popular forms of dipole antennas

Thus, the radiation efficiency is

$$\eta_e = \frac{R_r}{R_r + R_L} = \frac{0.0197}{0.0197 + 0.0126} = 60.99\% \quad \text{or} - 2.14\,\text{dB}$$

This efficiency is much lower than that at 1 GHz. Normally, electrically small antennas have low efficiencies and high reactances, which is why they are not popular in applications even though they are small in size. Sometimes we do employ short dipoles for reception where signal-to-noise ratio is more important than the radiation efficiency.

In addition to cylindrical dipoles, there are many other forms of dipoles developed for various applications. For example, as shown in Figure 5.4, the *biconical antenna* offers a much wider bandwidth than the conventional cylindrical dipole; a *bow-tie antenna* has a broad bandwidth and low profile; while a *sleeve dipole* can be nicely fed by a coaxial cable and exhibits a wide bandwidth. A λ/2 *folded dipole* may be viewed as a superposition of two half-wavelength dipoles with an input impedance of around 280 Ω, which is close to the standard characteristic impedance of 300 Ω for a two-wire transmission line. That is why it was widely used for TV reception in the good old days (in the 50s and 60s). More details about these and other forms of dipoles, as well as the effects of parameters such as the feeding gap, may be found in [2, 3, 4].

5.1.2 Monopoles and Image Theory

The monopole antenna is half of the dipole antenna, as shown in Figure 5.5. There are a lot of similarities between them, but there are also some differences. The best way to investigate the monopole is to utilize the image theory.

5.1.2.1 The Image Theory

The *image theory* states that if there is a current ***A/B/C*** above an infinite perfect conducting ground plane, the ground will act as a mirror to generate its image, ***A'/B'/C'***, as shown in Figure 5.6. ***A*** and ***B*** are the two basis currents; any other case, such as ***C***, may be viewed as a combination of these two currents. The field at any point above the ground plane is equivalent

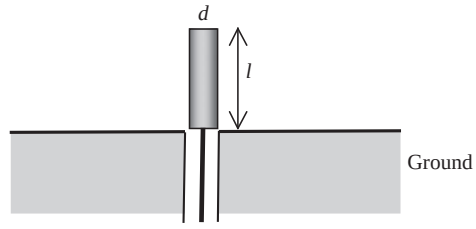


Figure 5.5 A monopole antenna with a coaxial feed line

to the field generated by the current ***A/B/C*** and its image ***A'/B'/C'*** without the presence of the ground plane.

The image current is of the same amplitude as the original source and its direction is determined by the boundary conditions, such as that stating that the tangential electric field must be zero at the boundary. Using the image theory, we can remove the ground plane and treat the problem as if it were in free space but had a pair of current sources.

5.1.2.2 Monopole Antennas

Applying the image theory to the monopole in Figure 5.5, we can see that it is equivalent to the case of ***B–B'*** in Figure 5.6, which is a dipole with length $2l$ in free space. The current distribution along the pole is the same as the dipole discussed earlier, thus the radiation pattern is the same above the ground plane and can be represented by Equation (5.7) with $0 \leq \theta \leq \pi/2$. Since the power is only radiated to the upper half space and the power to the lower half space is reflected back to the upper space, this results in an increased directivity. The directivity of a monopole is therefore twice that of its dipole counterpart – this can also be obtained using Equation (5.9) when the range of the integral is changed to $0 \leq \theta \leq \pi/2$.

The input impedance is changed as well. For a dipole, the voltage across the input points of the poles is $V - (-V) = 2V$, while for a monopole, the voltage is between the pole and the ground plane, which is V. The input currents for both the dipole and monopole are the same. Thus, the input impedance of the monopole is half that of its corresponding dipole. A summary of the characteristics of some monopoles is given in Table 5.2. Compared with the dipole, the monopole has the following advantages:

- the size is half of the corresponding dipole's;
- the directivity doubles that of its corresponding dipole;
- the input impedance is half of its corresponding dipole's.

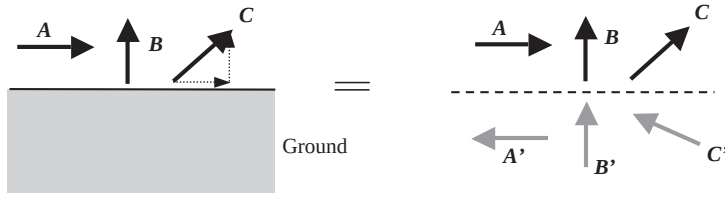
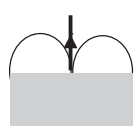
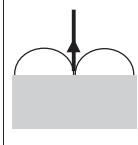
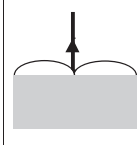
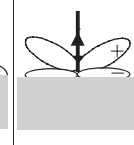


Figure 5.6 The image theory

Table 5.2 Summary of some monopole characteristics

Monopole length l	$\lambda/20$	$\lambda/4$	$\lambda/2$	3/4
Current distribution				
Radiation pattern				
Directivity	3.0 or 4.76 dBi	3.28 or 5.15 dBi	4.8 or 6.8 dBi	About 4.6
HPBW	45°	39°	23.5°	NA
Input impedance	R: very small ($\sim 1\Omega$) jX: capacitive	R: $\sim 37\ \Omega$ jX: $\sim 0\ \Omega$	R: very large jX: $\sim 0\ \Omega$ for thin dipole	R: $\sim 50\ \Omega$ jX: $\sim 0\ \Omega$ for thin dipole
Note	jX sensitive to the radius	$R+jX$ not sensitive to the radius	$R+jX$ sensitive to the radius	$R+jX$ sensitive to the radius

The quarter-wavelength monopole is a resonant antenna, like the half-wavelength dipole, its input impedance is about 37 Ω, which matches well with the 50 Ω standard transmission line (with $VSWR = 1.35 < 2$). All these good reasons have made the quarter-wavelength monopole one of the most popular antennas. We can find applications almost everywhere, from radio broadcasting towers to mobile phones.

Just like dipoles, there are many derivatives of monopole antennas developed to suit different applications. Some are shown in Figure 5.7. More detailed discussion will be presented in Chapter 8.

5.1.2.3 Effects of a Ground Plane

The above analysis and characteristics of monopoles are based on the assumption that there is an infinite perfect conducting ground plane. But in reality, we normally do not have an infinite ground plane or the ground plane is not a perfect conductor (like earth). What are the effects on the monopole?

All parameters of the monopole (radiation pattern, gain and input impedance are just some examples) may be affected by the ground plane.

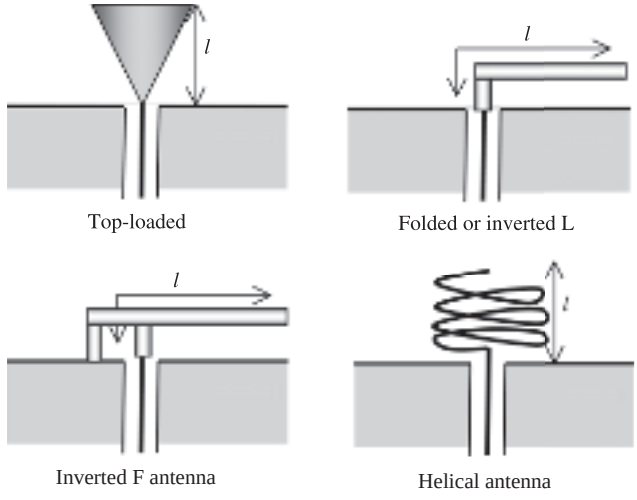
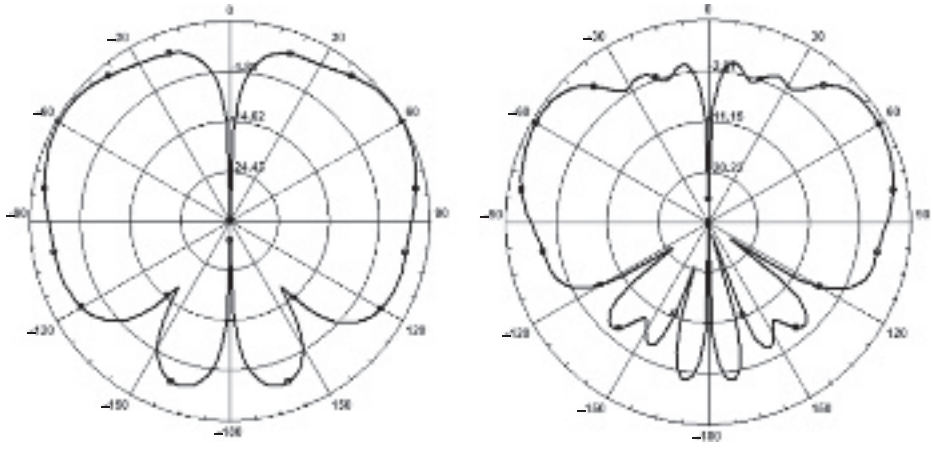


Figure 5.7 Some popular forms of monopole antennas

If the conducting ground plane is of limited size, the radiated power will leak to the lower half of the space, which means that the radiation pattern is changed. There may be side or even back lobes. The edge of the ground plane will diffract the waves, which results in many side lobes, as shown in Figure 5.8. The maximum angle is changed (tilted towards the sky) and the directivity is reduced. Also, the input impedance may be changed. If the ground plane is not large enough, it can act as a radiator rather than a ground plane. As a rule of thumb, the diameter of the ground plane should be at least one wavelength.

If the ground plane is very large but not made of a good conductor, all the antenna properties are affected, especially the directivity and gain (reduced); the angle to the maximum radiation



$Gain_{total}$ *(dB)*, Ground $R = \lambda$ E_{total} *(dB)*, Ground $R = 2 \times \lambda$

Figure 5.8 Effects of the ground plane on the radiation pattern of a monopole [3]

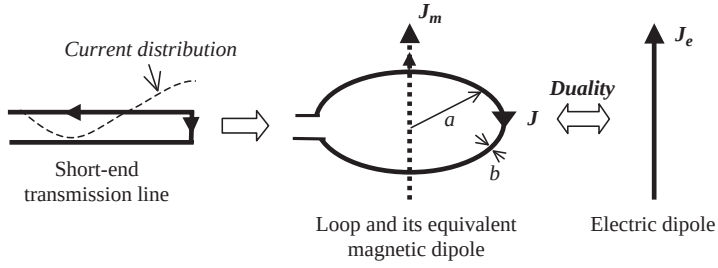


Figure 5.9 Transmission line to loop antenna and its corresponding dipoles

is also tilted towards the sky, although there are still no lobes in the lower half space. To improve the reflectivity of the ground, metal meshes are sometimes employed. A good discussion can be found in [2].

5.1.3 Loops and the Duality Principle

Loops are another simple and versatile wire-type antenna. They take many different configurations, which include circular, square, rectangular, triangular, elliptical and other shapes.

While the dipole is considered to be a configuration evolved from an open-end transmission line, the loop can be viewed as a configuration evolved from a short-end transmission line, as shown in Figure 5.9. Thus, for a small (electrically small) loop, the current is large and the voltage is small – this results in a small input impedance, which is very different from a short dipole whose impedance, more precisely the reactance, is very large. We could use the conventional method to obtain a loop's properties by finding its current distribution first, but this is time-consuming and cumbersome. Since we have already studied dipoles properly, we are going to use another method, the duality principle, to find the characteristics of a loop, which is easier and more straightforward.

5.1.3.1 Duality Principle

Duality means the state of combining two different things which are closely linked. In antennas, the *duality theory* means that it is possible to write the fields of one antenna from the field expressions of the other antenna by interchanging parameters.

From Chapters 1 and 4, we know that the first two Maxwell equations are

$$\begin{aligned} \nabla \times \boldsymbol{E}_1 &= -j\omega\mu_1 \boldsymbol{H}_1 \\ \nabla \times \boldsymbol{H}_1 &= \boldsymbol{J}_e + j\omega\varepsilon_1 \boldsymbol{E}_1 \end{aligned} \tag{5.14}$$

where $\boldsymbol{E}_1$ and $\boldsymbol{H}_1$ are the electric and magnetic fields generated by electric current density $\boldsymbol{J}_e$ in Medium 1 (ε_1 and μ_1).

Table 5.3 Duality relationship between System 1 and System 2

System 1 with electric current source	System 2 with magnetic current source
$\boldsymbol{J}_e$	$\boldsymbol{J}_m$
$\boldsymbol{E}_1$	$\boldsymbol{H}_2$
$\boldsymbol{H}_1$	$-\boldsymbol{E}_2$
ε_1	μ_2
μ_1	ε_2

Now suppose we have a fictitious magnetic current source with magnetic current density $\boldsymbol{J}_m$ in Medium 2 (ε_2 and μ_2); Maxwell's equations for this new scenario can be written as

$$\begin{aligned} \nabla \times \boldsymbol{H_2} &= j\omega\varepsilon_2 \boldsymbol{E_2} \\ \nabla \times \boldsymbol{E_2} &= -\boldsymbol{J}_m - j\omega\mu_2 \boldsymbol{H_2} \end{aligned} \tag{5.15}$$

where $\boldsymbol{E}_2$ and $\boldsymbol{H}_2$ are the electric and magnetic fields generated by $\boldsymbol{J}_m$ in Medium 2 (ε_y and μ_y).

Thus, System 1 with the electric current and System 2 with the magnetic current are duals and their parameters can be exchanged, as shown in Table 5.3. Note that the duality links not only the field parameters but also the material properties; there is also a sign change from the magnetic field to the electric field.

5.1.3.2 Small Loops

The implementation of the duality principle is simple. For small loops (circumference $C = 2\pi a < \lambda/3$), the current can be considered a constant. It has been found that a magnetic current $I_m \Delta l$ is equivalent to a small loop of radius a and constant electric current I_0 provided that

$$I_m \Delta l = j\omega\mu S I_0 \tag{5.16}$$

where $S = \pi a^2$ (area of the loop). This means that their radiated fields are identical. As shown in Figure 5.9, using the duality principle, it is possible to write the fields of the magnetic current from the field expressions of the electric current antenna by interchanging parameters. From Equations (4.6) and (4.7), the far field of a short electric current is given as

$$\begin{aligned} E_\theta &= \frac{jI\Delta l}{4\pi r}\eta\beta \sin\theta e^{-j\beta r} \\ H_\phi &= \frac{jI\Delta l}{4\pi r}\beta \sin\theta e^{-j\beta r} \\ E_r &= E_\phi = 0;\ H_r = H_\theta = 0 \end{aligned} \tag{5.17}$$

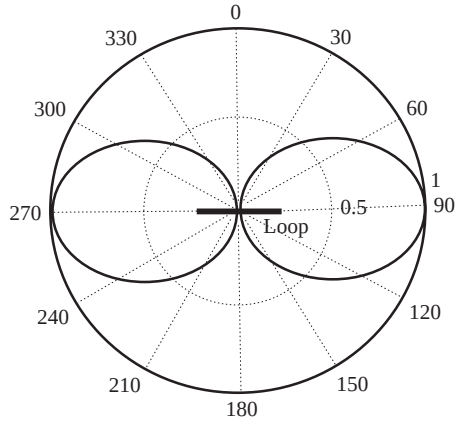


Figure 5.10 E_ϕ radiation pattern of a small loop as a function of θ (H-plane)

We use Table 5.3 to yield the far fields of an electrically small loop

$$\begin{aligned} H_\theta &= \frac{j I_m \Delta l}{4\pi r \cdot \eta} \beta \sin\theta e^{-j\beta r} \\ E_\phi &= -\frac{j I_m \Delta l}{4\pi r} \beta \sin\theta e^{-j\beta r} \\ H_r &= H_\phi = 0;\ E_r = E_\theta = 0 \end{aligned} \tag{5.18}$$

The radiation pattern looks the same as a short dipole, as shown in Figure 5.10, except that *the polarization is now E_ϕ (not E_θ, but still linearly polarized)*, which is in the plane of the loop. The maximum radiation occurs at $\theta = \pi/2$, which is also in the loop plane.

The directivity of a small loop is the same as that of a small dipole, which is 1.5 or 1.76 dBi.

Using the radiated field expressions in (5.18), the radiation resistance is found as

$$R_r = \frac{P_t}{I_0^2} = 20\pi^2(\beta a)^4 = 20\pi^2 \left(\frac{C}{\lambda}\right)^4 \tag{5.19}$$

where C is the circumference of the small loop. If it is not a circle, C is then the total length of the loop. Comparing Equation (5.19) with a dipole's radiation resistance, given in Equation (5.13), we can see that the loop radiation resistance is more sensitive to changes in the length and wavelength. For a loop and dipole of the same length, the radiation resistance of the loop antenna is much smaller.

The reactance of the small loop can be approximated by [2]

$$X_A = \omega\mu a \left(\ln\left(\frac{8a}{b}\right) - 2\right) + \frac{a}{b}\sqrt{\frac{\omega\mu}{2\sigma}} \tag{5.20}$$

where b is the diameter of the wire. This equation takes both the external inductive reactance (the first term) and internal reactance (the second term) of the loop conductor into account. For impedance matching, a capacitor in series is required for a small loop.

5.1.3.3 Loop Antenna: General Case

If a loop cannot be considered small, the current distribution cannot be regarded as constant. As a result, many properties of the loop are changed and results obtained for a constant current cannot be applied to real antennas. It has been shown that, when the circumference of the loop becomes comparable with the wavelength, its maximum radiation shifts to its axis ($\theta = 0$ and π), which is perpendicular to the plane of the loop [5]. This is very different from a small loop. There is no simple mathematical expression for the radiated field from such a loop antenna. Computer simulations are really the best choice if a 3D radiation pattern is required. The directivity of a circular loop as a function of the circumference in wavelengths (C/λ) for $a/b = 40$ and $\theta = 0$ is shown in Figure 5.11. The maximum directivity is about 4.5 dBi when the circumference is near 1.4λ. It should be pointed out that the directivity is fairly independent of the radius of the wire (b) for $C/\lambda < 1.4\lambda$.

The *one-wavelength loop* (the circumference $C = \lambda$) is commonly referred to as a resonant loop. The current in the loop is approximately $I_0 \cos\phi$ (maximum at the input) and is approximately equivalent to two $\lambda/2$ dipoles separated by $2a$ ($= \lambda/\pi$), as shown in Figure 5.12. E_θ, E_ϕ and the total radiation patterns of this antenna are plotted in Figure 5.13. E_θ is zero in the horizontal plane $\theta = \pi/2$ and in the vertical plane $\phi = 0, \pi$, while E_ϕ is small in the vertical plane $\phi = \pi/2, 3\pi/2$. The maximum is indeed shifted to the axis, which can be easily explained using the model in Figure 5.12. The directivity of the antenna is around 2.2 or 3.4 dBi (larger than for a $\lambda/2$ dipole). The input resistance is about 100 Ω, which is a reasonable value for matching to a standard 50 or 75 Ω transmission line. However, the input reactance in

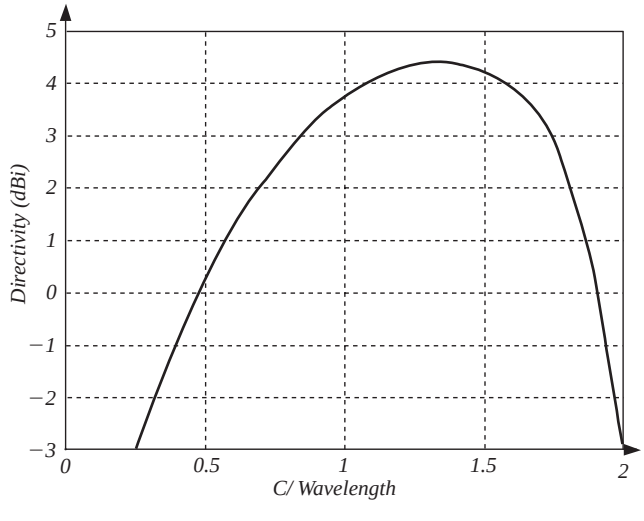


Figure 5.11 Directivity of a loop as a function of the normalized circumference

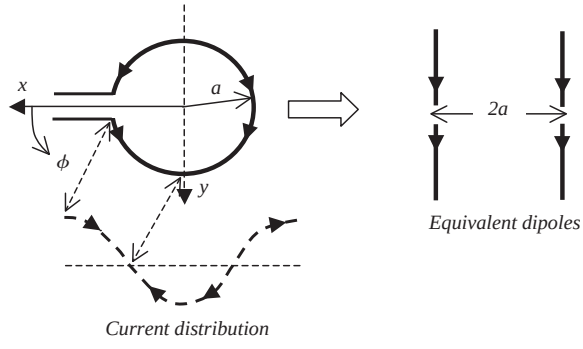


Figure 5.12 Current distribution in a resonant loop and its equivalent pair of dipoles

this case is actually about $-100\ \Omega$ (not sensitive to a change of radius). The resonance occurs when C is slightly larger than λ, as seen in Figure 5.14.

The input impedance of a loop with the circumference $C = 6$ cm (the corresponding frequency is 5 GHz) is shown in Figure 5.14. The input resistance reaches the first maximum at about $C/\lambda = 0.5$ (at 2.5 GHz, it is referred to as an anti-resonant loop), and the second maximum is at approximately $C/\lambda = 1.5$ (it is not at 9 GHz but near 8 GHz in Figure 5.14). It is interesting to note that the resistance curves for the loop and dipole (see Figure 5.2) are similar, and their reactance curves are also similar but with a reference position shift of about $\lambda/2$. This is not difficult to understand using the transmission line theory: the open and short points are shifted by $\lambda/2$ on a transmission line.

5.1.3.4 Discussion

The loop antenna has some very interesting features in terms of the input impedance and radiation pattern, as shown in Figures 5.14 and 5.15. When the loop size is changed from electrically small to about one wavelength, the total radiation is gradually changed from an omnidirectional pattern in the horizontal plane to an omnidirectional pattern in the vertical plane. If its size is further increased, the radiation pattern will no longer be omnidirectional but will have many lobes. Thus, the loop antenna radiation pattern is much more complicated than its dipole counterpart.

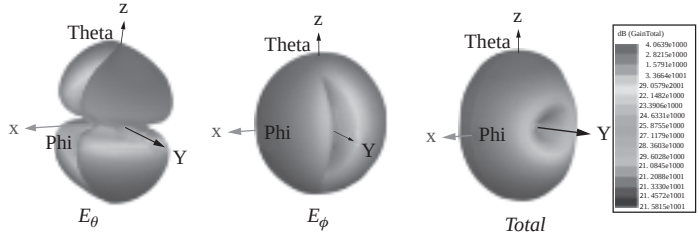


Figure 5.13 The radiation patterns of a loop with $C = \lambda$

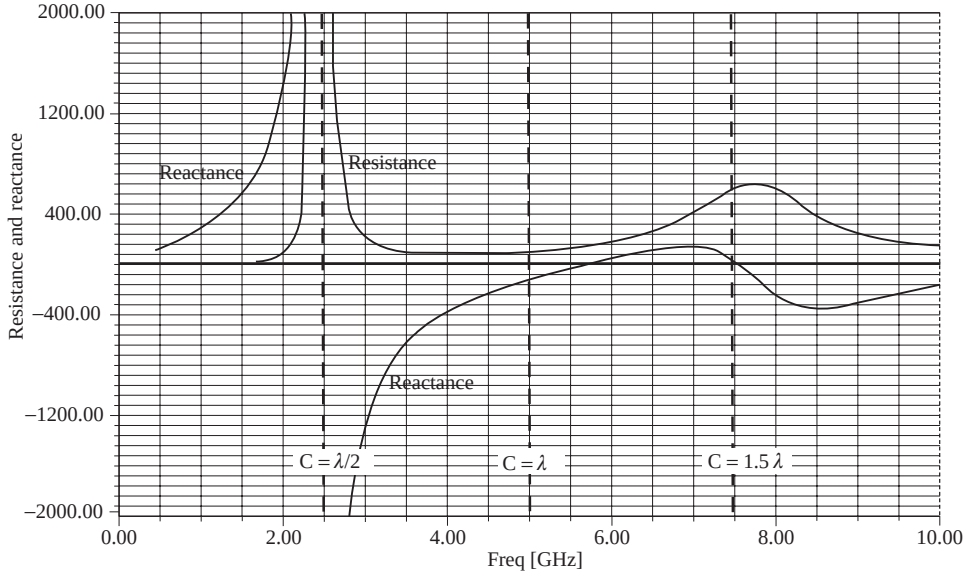


Figure 5.14 The input impedance of a loop with $C = 6$ cm

In addition to the circular loop, other forms of loop antennas have also been developed for practical applications. Since the analysis of polygonal (such as square, rectangular, triangular and rhombic) loops is much more complex, they have received much less attention.

Generally speaking, loop antennas have many attractive characteristics. They are of low profile and present a well-controlled radiation pattern. Balanced and unbalanced feeds are possible (coaxial cables can be used to feed the loop or half loop [2]). Loop antennas are usually employed for linear polarization although the shape is circular. Recently, it has been found that a loop antenna can also radiate circularly polarized waves if a gap is introduced on the loop [6, 7, 8].

A one-wavelength loop with a planar reflector has a unidirectional pattern with a relatively high directivity (near 9 dBi) and good input impedance, which is, of course, dependent on the distance to the reflector [3]. For mobile radio communications, the loop is often used in pagers but hardly used in mobile transceivers. This may be due to its high resistance and

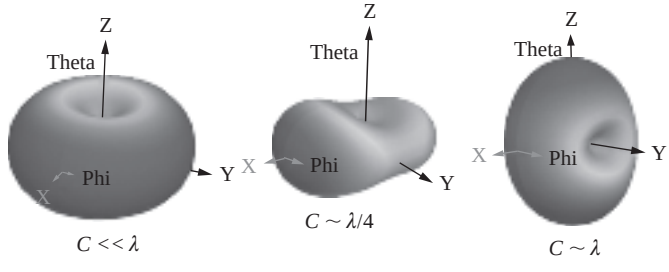


Figure 5.15 Total radiation patterns of loops of different sizes

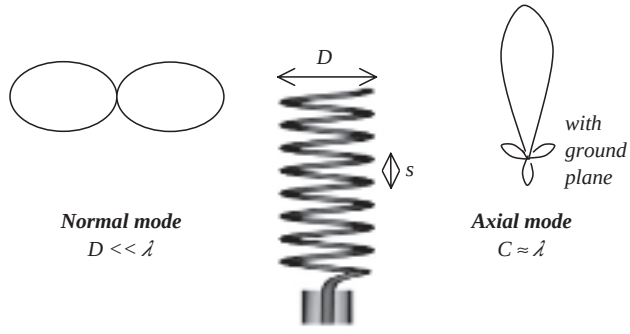


Figure 5.16 Helical antenna and the radiation patterns of two radiation modes

reactance, which make it difficult to match to standard 50 Ω transmission lines over moderate bandwidths. With a ground plane or body, the loop antenna should be an attractive candidate for mobile hand-portables. We expect to see more loop and derivative antennas being used in mobile wireless communication systems.

5.1.4 Helical Antennas

As indicated earlier, the helical antenna may be viewed as a derivative of the dipole or monopole antenna, but it can also be considered a derivative of a loop antenna. In this section we are going to introduce this special antenna, which has some very interesting and unique features, as Professor Kraus (who made a significant contribution to this subject) stated in [4], 'not only does the helix have a nearly uniform resistive input over a wide bandwidth but it also operates as a super-gain end-fire array over the same bandwidth! Furthermore, it is non-critical with respect to conductor size and turn spacing. It is easy to use in arrays because of almost negligible mutual impedance'.

Geometrically, as shown in Figure 5.16, a *helical antenna* consists of a conductor wound into a helical shape. It is a circularly/elliptically polarized antenna. A helix wound like a right-hand (clockwise) screw radiates or receives right-hand circularly polarized waves, whereas a helix wound like a left-hand (anti-clockwise) screw radiates or receives left-hand circularly polarized waves. Although *a helix can radiate in many modes*, the axial (end-fire) mode and the normal (broadside) mode are the ones of most interest. The following notation is used for the analysis:

$D =$ diameter of helix;
$s =$ spacing between turns;
$n =$ number of turns;
$C =$ circumference of helix $= \pi D$;
$L =$ length of one turn $= \sqrt{C^2 + s^2}$.

5.1.4.1 Normal Mode Helix

The normal (broadside) mode occurs when the diameter of the helix is much smaller than the wavelength ($D \ll \lambda$) and the total length is also smaller than the wavelength. It behaves more

like a dipole (or monopole with a ground plane) antenna. The radiation pattern is broadside to the helix axis. In this case, *the helix may be treated as the superposition of n elements, each consisting of a small loop of diameter D and a short dipole of length s*. The axes of the loop and dipoles coincide with the axis of the helix. The far field of a small loop is given in Equations (5.18) and (5.16) as

$$\begin{aligned} E_\phi &= \frac{\omega\mu S I_{in}}{4\pi r}\beta \sin\theta e^{-j\beta r} \\ E_r &= E_\theta = 0 \end{aligned} \tag{5.21}$$

The far field of a short dipole is shown in Equation (5.8) as

$$E_\theta \approx j\eta \frac{I_{in}}{4\pi r}(\beta s)\sin\theta e^{j\beta r} \tag{5.22}$$

The E_θ and E_ϕ components of dipoles and loops are 90 degrees out of phase. The combination of them gives a circularly or elliptically polarized wave. The axial ratio is

$$AR = \frac{|E_\theta|}{|E_\phi|} = \frac{\eta s}{\omega\mu S} = \frac{s}{2\pi f \cdot \sqrt{\varepsilon\mu}\cdot\pi(D/2)^2} = \frac{2\lambda s}{\pi^2 D^2} = \left(\frac{2s}{\lambda}\right)\Big/\left(\frac{\pi D}{\lambda}\right)^2 \tag{5.23}$$

When the circumference is equal to

$$C = \pi D = \sqrt{2s\lambda} \tag{5.24}$$

the axial ratio becomes unity and the radiation is circularly polarized. Otherwise, the radiation is elliptically polarized.

From Equations (5.21) and (5.22), we know that the radiation pattern is indeed a typical figure eight (8) in the vertical plane, as shown in Figure 5.16, and omnidirectional in the horizontal plane. The directivity is again about 1.5 and the HPBW is 90 degrees, similar to that of a short dipole or small loop.

The input impedance is very sensitive to changes of frequency; the bandwidth is therefore very narrow. In practice, the helix is normally used with a ground plane, the polarization is predominantly vertical and the radiation pattern is similar to that of a monopole. In this case, the radiation resistance can be approximately given by $(25.3\ ns/\lambda)^2$ [9]. These antennas tend to be inefficient radiators (depending on size and materials) and are typically used for mobile communications where reduced size is a critical factor.

5.1.4.2 Axial Mode Helix

The axial (end-fire) mode occurs when the circumference of the helix is comparable with the wavelength ($C = \pi \mathrm{D} \approx \lambda$) and the total length is much greater than the wavelength. This has made the helix an extremely popular circularly polarized broadband antenna at the VHF and UHF band frequencies. In this mode of operation, there is only one main lobe and its maximum is along the axis of the helix and, as shown in Figure 5.16 (with a ground plane), there may

be some side lobes. The recommended parameters for an optimum design to achieve circular polarization are:

- Normalized circumference: $3/4 < C/\lambda < 4/3$;
- Spacing: $s \approx \lambda/4$;
- Pitch angle: $12^\circ \leq \alpha = \tan^{-1}(s/C) \leq 15^0$;
- Number of turns: $n > 3$.

In contrast to the normal mode helix, which has a current almost uniform in phase over the antenna, the phase of the axial mode helix current shifts continuously along the helix like a traveling wave. Because the circumference is about one wavelength, the currents at opposite points on a turn are about 180degrees out of phase. This cancels out the current direction reversal introduced by the half turn. Thus, the radiation from opposite points on the helix is nearly in phase. This is essentially the same as the one-wavelength loop shown in Figure 5.12. It is therefore not surprising to have the maximum radiation along the axis.

The radiation pattern of the axial mode helix can be modeled using antenna array theory, to be discussed later in this chapter. Basically, each turn can be considered an element of the array with a radiation pattern of $\cos\theta$. The normalized total radiation pattern is:

$$E = A\cos\theta\frac{\sin[(n/2)\Psi]}{\sin[\Psi/2]} \tag{5.25}$$

where the first term is the normalization factor: $A = 1/n$ for ordinary end-fire radiation and $A = \sin(\pi/(2n))$ for Hansen–Woodyard (HW) end-fire radiation (*the axial mode helix was found to approximately satisfy the Hansen–Woodyard condition*, to be discussed in Section 5.3.2), which is larger than $1/n$ to reflect the increased directivity – it is achieved by increasing the phase change between element sources. The second term in the equation ($\cos\theta$) is the element radiation pattern and the last term represents the array factor of a uniform array of n elements. $\Psi = \beta(s\cos\theta - L/p)$ and p is the relative phase velocity given by [2, 4]

$$p = \frac{v}{c} = \begin{cases} \dfrac{L/\lambda}{s/\lambda + 1}; & \text{for ordinary end – fire radiation, } \Psi = -2\pi \\ \dfrac{L/\lambda}{s/\lambda + \dfrac{2n+1}{2n}}; & \text{for HW end – fire radiation, } \Psi = -2\pi - \pi/n \end{cases} \tag{5.26}$$

It has been found that the half-power beamwidth is roughly

$$HPBW \approx \frac{52^\circ\lambda\sqrt{\lambda}}{C\sqrt{ns}} \tag{5.27}$$

It is inversely proportional to C and $\sqrt{ns}$. The beamwidth between first nulls is about

$$FNBW \approx \frac{115^\circ\lambda\sqrt{\lambda}}{C\sqrt{ns}} \tag{5.28}$$

Using the link between the directivity and HPBW in Equation (4.14), we obtain the directivity:

$$D \approx 15C^2ns/\lambda^3 \tag{5.29}$$

This is proportional to C^2, n and s. Thus, the effects of n on directivity are significant when n is small but not significant when n is large (for example, if n is increased from 2 to 4, the directivity is doubled. If n is increased from 20 to 22, the increment is still 2, but the directivity has little change). If we take the minor lobes and the details of the pattern shape into account, a more realistic estimation is

$$D \approx 12C^2ns/\lambda^3 \tag{5.30}$$

The axial ratio is found to be

$$AR = (2n+1)/(2n) \tag{5.31}$$

The input impedance of an axial mode helix with a ground, which is greater than one wavelength, is almost resistive with values between 100 and 200 Ω. The estimated value within 20% accuracy is expressed by

$$R = 140C/\lambda \tag{5.32}$$

A suitable matching circuit is required by the antenna if it is to be connected to a 50 Ω transmission line. One way to bring the input impedance down to the desired value is to use the first 1/4 turn as an impedance transformer. A microstrip is an ideal structure for this purpose, since the ground plane may be the same as that of the antenna. The dielectric substrate height h is linked to the strip width w and feed line characteristic impedance Z_0 by [4]

$$h = \frac{w}{\frac{377}{\sqrt{\varepsilon_r}Z_0} - 2} \tag{5.33}$$

where ε_r is the relative permittivity of the substrate. This improved-matching modification may result in bandwidth reduction as a trade-off.

Example 5.2: Axial helix. Design a circularly polarized helix antenna of an end-fire radiation pattern with a directivity of 13 dBi. Find out its input impedance, *HPBW*, *AR* and radiation pattern.

Solution:
Using the recommended design parameters for an axial mode helix, we can choose: $s = \lambda/4$ and $C = \lambda$, which gives us the pitch angle: $12^\circ \leq \alpha = \tan^{-1}(s/C) = 14.0362^\circ \leq 15^\circ$.

The required directivity is

$$D = 13\,dBi = 20 \approx 12C^2ns/\lambda^3$$

thus, $n = 6.667 \approx 7$ turns. Now we can use the equations above to obtain:

$$\text{input impedance: } R = 140C/\lambda = 140\Omega;$$

$$\text{half-power beamwidth: } HPBW \approx \frac{52^\circ\lambda\sqrt{\lambda}}{C\sqrt{ns}} = 39.3^\circ;$$

$$\text{axial ratio: } AR = (2n+1)/(2n) = 1.0714.$$

To calculate the radiation pattern, we first find $L = \sqrt{C^2 + s^2} = 1.03\,\lambda$; and

$$p = \frac{v}{c} = \begin{cases} \dfrac{L/\lambda}{s/\lambda + 1} = 0.8295; & \text{for normal end – fire radiation, } \Psi = -2\pi \\ \dfrac{L/\lambda}{s/\lambda + \dfrac{2n+1}{2n}} = 0.7844; & \text{for HW end – fire radiation, } \Psi = -2\pi - \pi/n \end{cases}$$

Thus, for the normal end-fire radiation:

$$\Psi_1 = \beta(s\cos\theta - L/p) = 2\pi(0.25\cos\theta - 1.03/0.8295) = 1.5708\cos\theta - 7.8019$$

$$E_1 = \frac{1}{7}\cos\theta\frac{\sin[3.5\Psi_1]}{\sin[0.5\Psi_1]}$$

and for HW end-fire radiation, which can be achieved by this helix antenna:

$$\Psi_2 = \beta(s\cos\theta - L/p) = 2\pi(0.25\cos\theta - 1.03/0.7844) = 1.5708\cos\theta - 8.2505$$

$$E_2 = \sin\left(\frac{\pi}{14}\right)\cos\theta\frac{\sin[3.5\Psi_2]}{\sin[0.5\Psi_2]}$$

Both patterns (E_1: dashed line for normal end-fire array and E_2: solid line for the helix) are plotted in Figure 5.17. We can see that an HW design can indeed provide an increased directivity (by a factor of 1.115 or 0.918 dB in this case). The logarithmic presentation in the rectangular plot shows more and clearer details of the radiation than that of the polar plot in linear scale. It is also interesting to note that the increased directivity pattern E_2 has a sharper main beam but higher side lobe levels than that of the ordinary end-fire pattern E_1.

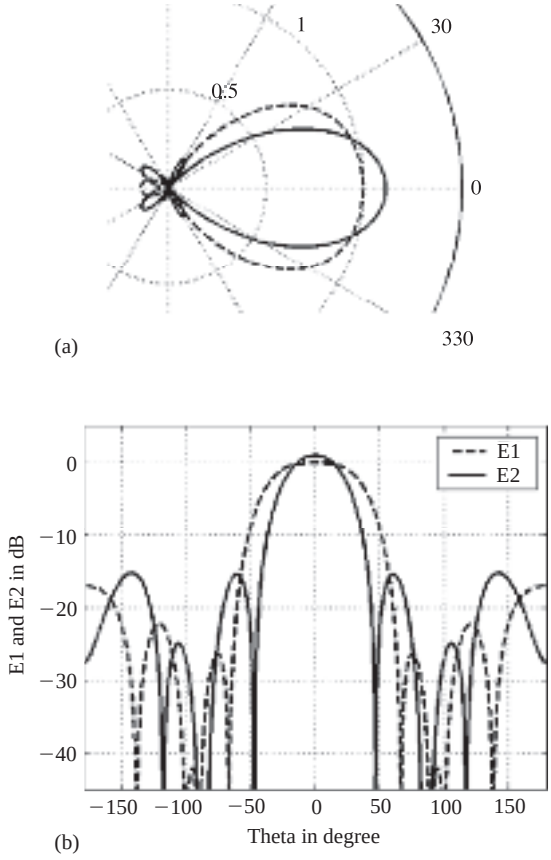


Figure 5.17 Radiated field patterns of two different antennas (a) the polar plot of the E_1 (dashed line) and E_2 (solid line) in linear scale; (b) E_1 and E_2 field patterns in dB scale

5.1.5 Yagi–Uda Antennas

The *Yagi–Uda antenna* (also known as a *Yagi*) is another popular type of end-fire antenna widely used in the VHF and UHF bands (30 MHz to 3 GHz) because of its simplicity, low cost and relatively high gain. The most noticeable application is for home TV reception and these can be found on the rooftops of houses. A typical one is shown in Figure 5.18.

Yagi and Uda were two Japanese professors who invented and studied this antenna in the 1920s. S. Uda made the first Yagi–Uda antenna and published the results in Japanese in 1926 and 1927, and the design was further developed and published in English by his colleague Professor Yagi a year later [10]. Since then a significant of amount of work has been done theoretically and experimentally. A lot of data and results are available in the public domain.

The main feature of this type of antenna is that it consists of three different elements: the driven element, reflector and director, as shown in Figure 5.19. Some people consider the Yagi–Uda antenna an array, since it has more than one element. However, it has just one

Figure 5.18 A Yagi–Uda TV reception antenna

active element and feed port; all the other elements (the reflector and directors) are parasitic. Thus, some people consider it an element antenna rather than an antenna array. The main characteristics and design recommendations of these elements can be summarized as follows:

- **The driven element (feeder):** the very heart of the antenna. It determines the polarization and central frequency of the antenna. For a dipole, the recommended length is about 0.47λ to ensure a good input impedance to a 50 Ω feed line.
- **The reflector:** normally slightly longer than the driven resonant element to force the radiated energy towards the front. It exhibits an inductive reactance. It has been found that there is not much improvement by adding more reflectors to the antenna, thus there is only one reflector. The optimum spacing between the reflector and the driven element is between 0.15 and 0.25 wavelengths. The length of the reflector has a large effect on the front-to-back ratio and antenna input impedance.
- **The directors:** usually 10 to 20% shorter than the resonant driven element and appear to direct the radiation towards the front. They are of capacitive reactance. The director to

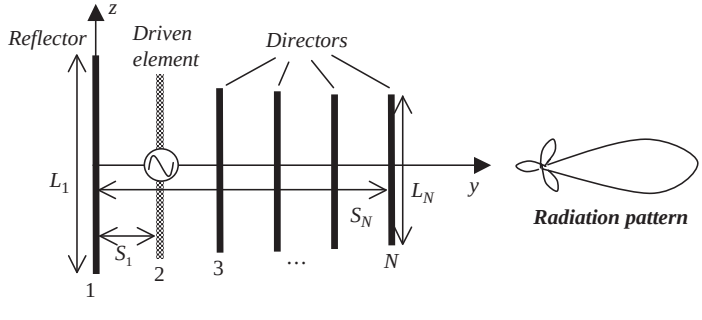


Figure 5.19 Configuration of a Yagi–Uda antenna and its radiation pattern

director spacing is typically 0.25 to 0.35 wavelengths, with larger spacing for longer arrays and smaller spacing for shorter arrays. The number of directors determines the maximum achievable directivity and gain.

5.1.5.1 Operational Principle

The special configuration (long reflector and short directors) has made the Yagi–Uda antenna radiate as an end-fire antenna. The simplest three-element Yagi–Uda antenna (just one director) already shows an acceptable end-fire antenna pattern. The radiation towards the back seems to be blocked/reflected by the longer element, but not just by the reflector; the reflector and director produce push-and-pull effects on the radiation. Induced currents are generated on the parasitic elements and form a traveling wave structure at the desired frequency. The performance is determined by the current distribution in each element and the phase velocity of the traveling wave.

5.1.5.2 Current Distribution

The current distribution on the driven element is determined by its length, frequency and interaction/coupling with nearby elements (mainly the reflector and first director), while the current distribution in parasitic elements is governed by the boundary condition: the total tangential electric field must be zero on the conducting surface. This results in induced currents and they may be viewed as the second sources of the radiation. Analytical and numerical methods have been employed to obtain the current distribution along each element [2]. The results show that the current distribution on each element is similar to that of a dipole. As expected, the dominant current is on the driven element; the reflector and the first director carry less current, and the currents on other directors are further reduced and they appear to be of similar amplitude, which is typically less than 40% of that of the driven element.

5.1.5.3 Radiation Pattern

Once the current is known, the total radiated field can be obtained using Equation (4.3), which may be difficult and numerical computations may be required.

For a dipole Yagi–Uda antenna, the radiation from element n is given by Equation (5.4)

$$E(\theta)_n \approx j\eta \frac{I_n e^{j\beta r}}{2\pi r} \left(\frac{\cos(\beta l_n \cos\theta) - \cos(\beta l_n)}{\sin\theta} \right) \tag{5.34}$$

where I_n is the maximum current and l_n is half the length of the nth dipole. Thus, the total radiation pattern is the field superposition from all the elements and may be expressed as

$$E(\theta) \approx j\eta \frac{e^{j\beta r}}{2\pi r} \sum_{n=1}^{N} I_n \left(\frac{\cos(\beta l_n \cos\theta) - \cos(\beta l_n)}{\sin\theta} \right) \exp(j\beta S_{n-1} \cos\theta) \tag{5.35}$$

where r is the center of the reflector to the observation point and spacing $S_0 = 0$. It is apparent that each element length and spacing, weighted by its maximum current, affects the total

radiation. This approach is the same as the one we are going to use for analyzing antenna arrays later.

One important figure of merit in a Yagi–Uda antenna is the *front-to-back ratio* of the pattern. It has been found that this is very sensitive to the spacing of the director. It varies from trough to peak and from peak to trough repetitively as a function of the spacing [2].

5.1.5.4 Directivity and the Boom

The directivity can be obtained using Equation (4.11), as we did for dipoles. A simpler estimation of the maximum directivity of a Yagi–Uda antenna is proposed by us as

$$D = 3.28N \tag{5.36}$$

The coefficient 3.28 results from doubling the directivity (1.64) of a half-wave dipole. Since there are N elements, the maximum is obtained when they are combined constructively as $3.28N$. The reason for introducing the factor of 2 is that the radiation pattern is now unidirectional end-fire. The radiation is redirected to just half of the space by the reflector and directors (very little to the other half), which is somewhat similar to the effect of a conducting ground plane. For the simplest three-element Yagi–Uda antenna, $N = 3$, thus $D = 9.84 = 9.93$ dBi $= 7.78$ dBd. When the number is doubled to $N = 6$, 3 dB more gain can be obtained ($D = 12.93$ dBi $= 10.78$ dBd). However, if three more directors are added to the antenna, the improvement in the directivity is just 1.76 dB. Just like the helical antenna, *the effect of the number of elements on the directivity is significant when N is small, but not significant when N is large*.

There is a misperception on how the directivity is linked to the length of the boom. Some people think the directivity is proportional to the length of the boom and not the number of elements. The reality is that, as indicated by Equation (5.36), the number of properly placed elements determines the directivity. Of course, the more elements, the longer the boom, and hence the larger the directivity. But the point is that the directivity is determined by the number of elements, not the length of the boom.

Some characteristics of the antenna may be affected by the boom if it is made of metal, for example the input impedance. The boom should be insulated from the elements, otherwise compensation is required. Since the orientation of the boom is orthogonal to the antenna elements, the effects on the radiation pattern should be small.

5.1.5.5 Input Impedance

The input impedance is very sensitive to the spacing of the nearest two elements and their lengths. *As a rule of thumb, the closer the spacing, the smaller the input impedance.* Coupling reduces the impedance. If a half-wavelength dipole is used as the driven element, its impedance can be reduced from 73 Ω to as little as 12 Ω when the spacing drops down to 0.1λ. This may be a problem if the antenna is to be matched with a 50 Ω or even a 75 Ω transmission line. A simple solution in practice is to replace a half-wavelength dipole by a folded dipole, which has a typical impedance of 280 Ω.

5.1.5.6 Antenna Design

Generally speaking, Yagi–Uda antennas have low impedance and relatively narrow bandwidth (a few percent). Improvement in both can be achieved at the expense of other parameters (such as directivity and side lobes). The length and diameters of the elements as well as their respective spacing determine the optimum characteristics. Since there are so many variables, the optimum design of a Yagi–Uda antenna is a very complicated task. Extensive and comprehensive experimental and theoretical investigations into Yagi–Uda antennas have been conducted. The most well-known experimental study was conducted by Viezbicke at NBS (now NIST, National Institute of Standards and Technology) [11]. He has produced a lot of data and results. Some of these are summarized in Table 5.4, along with our estimated maximum directivity. These data are extremely useful for practical antenna designs.

It should be pointed out that the element diameter is fixed at 0.0085λ in Table 5.4. If this condition cannot be met in reality, slight changes to the element length should be made and relevant design curves for this purpose were produced in [11], with some reproduced in [2, 12].

These results were obtained experimentally. Analytical and numerical investigations have shown that further improvements could be achieved. For example, Chen and Cheng [13, 14] used a perturbational technique to vary each spacing and element length and showed that the directivity of a six-element Yagi–Uda antenna was increased from the initial value (10.92 dBi) to 12.89 dBi, which is very close to our estimated maximum directivity of 12.93 dBi.

Some optimization methods, such as genetic algorithm and particle swarm, have recently been employed to optimize and miniaturize the design [15, 16, 17]; good and interesting results

Table 5.4 Optimized elements for Yagi–Uda antennas (the normalized diameter $d/\lambda = 0.0085$, spacing $S_1 = 0.2\lambda$)

Boom length/λ	0.4	0.8	1.2	2.2	4.2	Note
L_1/λ	0.482	0.482	0.482	0.482	0.475	Reflector
L_2/λ		$\lambda/2$ folded dipole ~ 0.47				Driven element
L_3/λ	0.442	0.428	0.428	0.432	0.424	Director
L_4/λ		0.423	0.420	0.415	0.424	
L_5/λ		0.428	0.420	0.407	0.420	
L_6/λ			0.428	0.398	0.407	
L_7/λ				0.390	0.403	
L_8/λ				0.390	0.398	
L_9/λ				0.390	0.394	
L_{10}/λ				0.390	0.390	
L_{11}/λ				0.398	0.390	
L_{12}/λ				0.407	0.390	
L_{13}/λ					0.390	
L_{14}/λ					0.390	
L_{15}/λ					0.390	
Spacing/λ	0.20	0.20	0.25	0.20	0.308	Between directors
D in dBd	7.1	9.2	10.2	12.25	14.2	Measured
D in dBi	9.2	11.3	12.3	14.35	16.3	Measured
Estimated D_{max} *in dBi*	9.93	12.14	12.93	15.95	16.91	Equation (5.35)

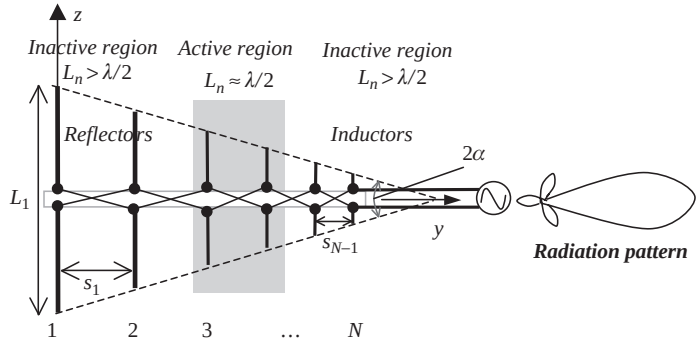


Figure 5.20 The configuration of a log-periodic antenna and its radiation pattern

have been obtained. It has been shown that optimum gains may be obtained with very small spacings. This is a very attractive feature for modern mobile communication systems, where size is particularly important.

In addition to the dipole, other antenna types (especially loops) have also been used to make Yagi–Uda antennas for various applications.

5.1.6 Log-Periodic Antennas and Frequency-Independent Antennas

A very similar configuration to the Yagi–Uda antenna is the *log-periodic antenna*, as shown in Figure 5.20. It produces a similar end-fire radiation pattern and directivity (typically between 7 and 15 dBi) to the Yagi–Uda and, also like the Yagi–Uda, is widely used in the VHF and UHF bands. However, there are two major differences between them:

- **Bandwidth:** the log-periodic antenna has a much wider bandwidth than the Yagi–Uda.
- **Feeder:** each element of the log-periodic antenna is connected to the source and can be seen as a feed of the antenna (i.e. each element is active) whilst there is only one feed (driven element) and all the others are parasitic elements in the Yagi–Uda antenna.

5.1.6.1 Operational Principle of Log-Periodic Antennas

As seen in Figure 5.20, the log-periodic dipole antenna (LPDA) consists of many dipoles of different lengths. The antenna is divided into the so-called *active region* and *inactive regions*. The role of a specific dipole element is linked to the operating frequency: if its length, L, is around half of the wavelength, it is an active dipole and within the active region; if its length is greater than half the wavelength, it is in an inactive region and acts as a reflector; if its length is smaller than half the wavelength, it is also in an inactive region but acts as a director, which is very similar to the Yagi–Uda antenna. The difference is that the driven element shifts with the frequency – this is why this antenna can offer a much wider bandwidth than the Yagi–Uda. A traveling wave can also be formed in the antenna. The highest frequency is basically determined by the shortest dipole length (L_N) while the lowest frequency is determined by the longest dipole length (L_1).

The reason for this antenna being called a *log-periodic antenna* is that its input impedance is a periodic function of the logarithm of the frequency. Other parameters that undergo similar variations include the radiation pattern, directivity and beamwidth.

5.1.6.2 Antenna Design

The geometrical dimensions of the log-periodic antenna follow some pattern and condition, which is another difference from the Yagi–Uda antenna. For analysis, the following notation is used:

L_n = the length of element n, and $n = 1, 2, \ldots, N$;
s_n = the spacing between elements n and $(n + 1)$;
d_n = the diameter of element n;
g_n = the gap between the poles of element n.

They are related to the *scaling factor*:

$$\tau = \frac{L_2}{L_1} = \frac{L_{n+1}}{L_n} = \frac{s_{n+1}}{s_n} = \frac{d_{n+1}}{d_n} = \frac{g_{n+1}}{g_n} < 1 \tag{5.37}$$

and the *spacing factor*:

$$\sigma = \frac{s_1}{2L_1} = \frac{s_n}{2L_n} < 1 \tag{5.38}$$

As shown in Figure 5.20, two straight lines through the dipole ends form an angle 2α, which is a characteristic of the frequency-independent structure. The angle α is called the *apex angle* of the log-periodic antenna, which is a key design parameter and can be found as

$$\alpha = \tan^{-1}\left(\frac{(L_n - L_{n+1})}{2s_n}\right) = \tan^{-1}\left(\frac{L_n(1-\tau)}{2s_n}\right) = \tan^{-1}\left(\frac{(1-\tau)}{4\sigma}\right) \tag{5.39}$$

These relations hold true for any n. From the operational principle of the antenna and the frequency point of view, Equation (5.37) corresponds to:

$$\tau = \frac{L_{n+1}}{L_n} = \frac{f_n}{f_{n+1}} \tag{5.40}$$

Taking the logarithm of both sides, we have

$$\log f_{n+1} = \log f_n - \log \tau \tag{5.41}$$

This means that the resonant frequency in log scale is increased by every $|\log \tau|$. Thus, the performance of the antenna is periodic in a logarithmic fashion, hence the name log-periodic antenna, as mentioned earlier.

This seems to have too many variables. In fact, *there are only three independent variables for the log-periodic antenna design*. These three parameters, which can be chosen from the

Table 5.5 Optimum design data for log-periodic antenna

Directivity/dBi	Scaling factor τ	Spacing factor σ	Apex angle α
7	0.782	0.138	21.55°
7.5	0.824	0.146	16.77°
8	0.865	0.157	12.13°
8.5	0.892	0.165	9.29°
9	0.918	0.169	6.91°
9.5	0.935	0.174	5.33°
10	0.943	0.179	4.55°
10.5	0.957	0.182	3.38°
11	0.964	0.185	2.79°

directivity, length of the antenna, apex angle, the upper frequency and the lower frequency, should come with the design specifications. Extensive investigations have been conducted and some optimum designs have been obtained. The most noticeable work was carried out by Carrel [18] and a correction was later made by Butson and Thompson in [19]. Computed contours of constant directivity (gain) versus σ and τ were given in these references and [2, 3, 12] (but some references have not incorporated the correction from [19]). A summary of the optimum design data (with correction) is produced in Table 5.5, which can be used to aid antenna design. It is apparent that, the higher the directivity, the larger the scaling factor and spacing factor, but the smaller the apex angle.

Another important aspect of the design is the antenna input impedance, which can be tuned by changing the diameter d of the element and the feeding gap $g$between the two poles.

$$g = d\cosh\left(Z_0/120\right) \tag{5.42}$$

where Z_0 is the characteristic impedance of the feed line to be connected (the desired impedance). More details can be found in [2, 18].

In practice, the most likely scenario is that the frequency range is given from $f_{\min}$ to $f_{\max}$; the following equations may be employed for design:

$$L_1 \geq \frac{\lambda_{\max}}{2} = \frac{c}{f_{\min}};\ L_N \leq \frac{\lambda_{\min}}{2} = \frac{c}{f_{\max}} \tag{5.43}$$

and

$$\frac{f_{\min}}{f_{\max}} = \frac{L_N}{L_1} = \tau\frac{L_{N-1}}{L_1} = \tau^{N-1} \tag{5.44}$$

Another parameter (such as the directivity or the length of the antenna) is required to produce an optimized design.

Once the geometrical dimensions are obtained, it is desirable to find the radiation pattern. Unfortunately, there is no analytical solution; the best approach is to use some numerical methods or software to complete this task. More discussion on this subject will be presented in the next chapter.

Example 5.3: Log-periodic antenna design. Design a log-periodic dipole antenna to cover all UHF TV channels, which is from 470 MHz for channel 14 to 890 MHz for channel 83. Each channel has a bandwidth of 6 MHz. The desired directivity is 8 dBi.

Solution:

The given three parameters are: $f_{\mathbf{min}} = 470\,\text{MHz},\ f_{\mathbf{max}} = 890\,\text{MHz},$ and $D = 8\,\text{dBi}$.

Since the desired directivity is 8 dBi, from Table 5.5, we can see that, for the optimum design, the scaling factor $\tau = 0.865$, the spacing factor $\sigma = 0.157$, and the apex angle $\alpha = 12.13^{\circ}$. The latter can also be obtained using Equation (5.39).

Because the frequency range is known, using Equation (5.44) yields

$$N = \log\left(\frac{f_{\min}}{f_{\max}}\right) \Big/ \log(\tau) + 1 = \log\left(\frac{470}{890}\right) \Big/ \log(0.865) + 1 = 4.40 + 1 = 5.40 \approx 6$$

That means at least six elements are required. To be on the safe side we should use seven or even eight elements to be sure the desired directivity will be achieved.

If $N = 8$, we can afford to start from a lower frequency, say 400 MHz, thus

$$L_1 = \frac{c}{f_{\mathbf{min}}} = \frac{300}{400} = 0.75\,(\text{m}),$$

and $L_2 = \tau L_1 = 0.865 * 0.75 = 0.6487\,(\text{m}), \ldots, L_8 = 0.2718 < \dfrac{c}{f_{\mathbf{max}}} = \dfrac{300}{890} = 0.3371\,(\text{m}).$

The spacing can be obtained using (5.38), that is

$$s_n = 2L_n\sigma = 0.314L_n$$

$n = 1$ to 7. ($s_1 = 0.2355; \ldots, s_7 = 0.0986$)

The total length of the antenna is

$$L = \sum_{n}^{7} s_n = 1.1142\,(\text{m})$$

this seems to be a reasonable length in real life.

In theory, the dipole diameter and feed gap at an element should also be scaled, as indicated in Equation (5.37). But in practice, this is sometimes hard to achieve. Because these two parameters do not affect the radiation pattern or directivity, we may just use the same or a few different metal poles/tubes to make the antenna. The gap is linked to how each element is fed and Equation (5.42) may be used to tune the impedance. It is not uncommon to use a constant gap, which means a matching section/circuit may be required if the antenna is not well matched with the feed line.

It should be pointed out that for the same number of elements, the Yagi–Uda antenna can produce a higher directivity than its log-periodic counterpart. For example, in Example 5.3, a six-element LPDA can offer 8 dBi gain, but a six-element Yagi–Uda antenna can achieve a gain of more than 12 dBi. This is due to the fact that the Yagi–Uda is a narrowband antenna and all elements are designed to contribute to the radiation constructively at the desired frequency,

while the log-periodic antenna is a broadband antenna and elements within the inactive regions may not contribute to a specific frequency as efficiently and constructively as in a Yagi–Uda antenna. There is a trade-off between the directivity and bandwidth for fixed N. Question 5.10 will reinforce this concept.

In addition to the log-periodic dipole antenna, there are other forms of log-periodic antennas, such as log-periodic loop antennas. The operation and design principle is the same.

5.1.6.3 Frequency-Independent Antennas

If the characteristics (such as the impedance and radiation pattern) of an antenna are not a function of frequency, the antenna is called a *frequency-independent antenna*. There are basically three methods for constructing such an antenna.

Scaling

If the antenna structure is built with a scaling factor τ, like the log-periodic antenna, its properties will be the same at f and τf. By making the scaling factor close to 1, the properties of the antenna at any frequency will be the same. In practice, even with τ not very close to 1, good frequency-independent characteristics are observed. The LPDA is a good example.

Angle Conditions

If the antenna shape is specified only in terms of angle, the impedance and pattern properties of the antenna will be frequency independent. Typical examples include various (wire, conical and planar) spiral antennas, as shown in Figure 5.21 [2, 3, 4, 12]. Mathematically, the equation for a logarithmic wire spiral can be expressed as

$$r = r_0 a^{\phi} \tag{5.45}$$

where

r = radial distance to a point on the spiral;
r_0 = the radius for $\phi = 0$, determining the upper frequency bound;
ϕ = angle with respect to the x-axis;
a = a constant controlling the flaring rate of the spiral. When $a > 1$, it is right-handed, while when $0 < a < 1$, it is left-handed.

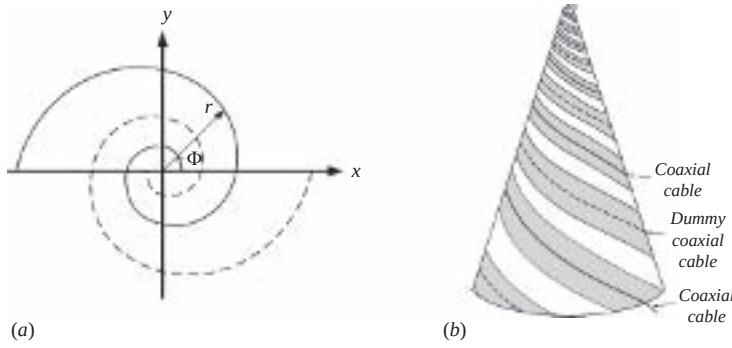


Figure 5.21 Spiral antennas (a) the wire type; (b) the conical type

A second wire spiral, which is identical in form to the first one with a rotation of π (now $\phi \geq \pi$), can be generated by

$$r = r_0 a^{\phi - \pi} \tag{5.46}$$

These two spirals form a circularly polarized frequency-independent antenna. The impedance, pattern and polarization remain nearly constant over a wide range of frequencies. The typical value of the *expansion ratio* $a^{2\pi}$ (the radius increase over one turn of the spiral) is 4. Experimental results suggest that a spiral of one and a half turns is about the optimum. Since the lower frequency bound is determined by the maximum radius, the bandwidth of a typical spiral is $r_0 a^{3\pi} : r_0 \approx 8 : 1$.

The input impedance is determined by the geometrical dimensions, especially around the antenna feeding. A typical value is between 120 and 180 Ω.

The radiation pattern is bidirectional with two wide beams broadside to the plane of the spiral, which is approximately $\cos\theta$ when the z-axis is normal to the plane of the antenna.

Self-complementary

If an antenna structure is identical to its complementary structure, its input impedance is frequency independent. This will be proved later in this chapter using *Babinet's principle*: the product of the input impedances of the original and the complementary structures is a constant = $\eta^2/4$ ($\eta \approx 377$ Ω in free space). The impedances for both structures without truncation should be identical, thus the input impedance is $\eta/2$ (≈ 188 Ω in free space). Two examples are shown in Figure 5.22.

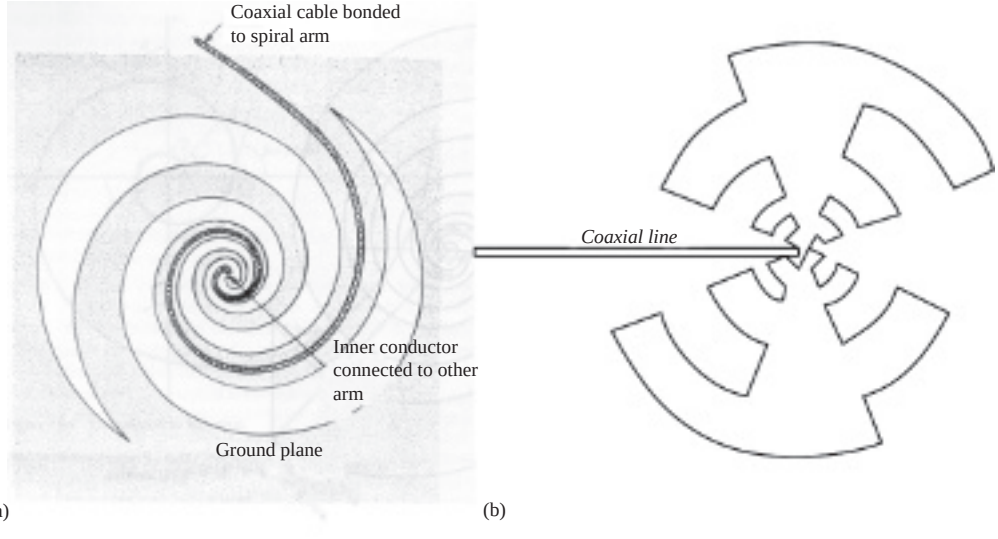


Figure 5.22 Two examples of self-complementary antennas (a) planar spiral; (b) log-periodic toothed planar antenna. (Reproduced by permission of John Wiley & Sons, Inc.)

In theory, these frequency-independent antennas should be infinitely large to cover the whole spectrum. In reality, the size has to be finite and therefore they have to be truncated to cover just the desired frequency range. The truncation has to be done carefully in order to minimize the effects on antenna performance. This means that the current has to be attenuated and negligible at the point of truncation – this is why the tail of a planar spiral antenna has a tapered truncation, as shown in Figure 5.22(a).

5.2 Aperture-Type Antennas

There is another group of antennas that are not made of metal wires but plates to form certain configurations that radiate/receive EM energy in an efficient and desired manner – the *aperture-type antennas*. They are often used for higher frequency applications than wire-type antennas. Typical examples include horn antennas and reflector antennas, as shown in Figure 5.23, which will be discussed in this section. Although we can still make use of the conventional approach to analyze this type of antenna, i.e. to obtain the current distribution first and then calculate the radiated field and input impedance, a new method, which is particularly suitable for analyzing aperture-type antennas, will be introduced in this section.

5.2.1 Fourier Transforms and the Radiated Field

It is well known that the Fourier transform is a very useful tool to convert a signal from the time domain, $x(t)$, to the frequency domain, $X(f)$. The mathematical expression is

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{j2\pi f t} dt \tag{5.47}$$

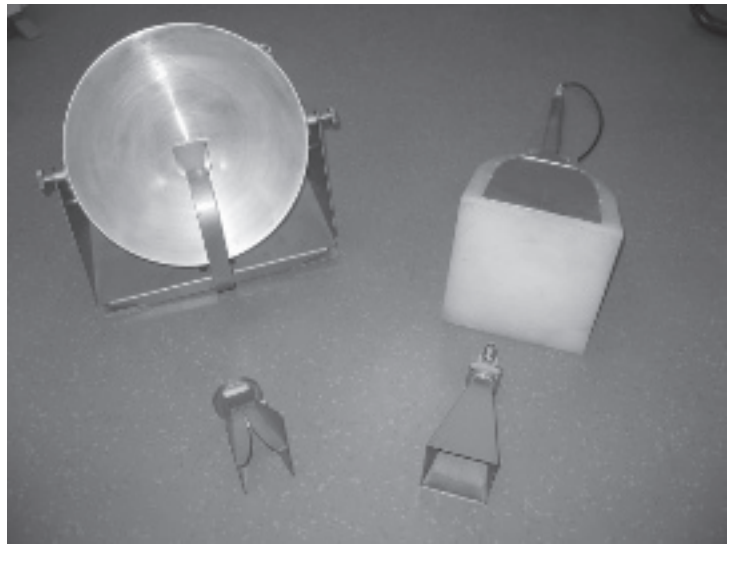


Figure 5.23 Some aperture-type antennas (reflector, TEM horn, double-ridged horn and pyramidal horn)

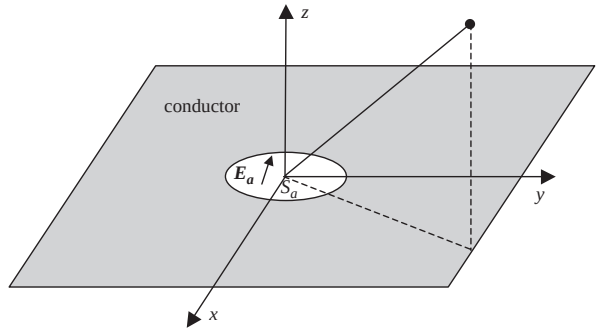


Figure 5.24 Radiation from an aperture source in the $z = 0$ plane

Now let us deal with the radiation from an aperture source S_a located in the $z = 0$ plane. As shown in Figure 5.24, it is assumed that the electric field on this aperture is known as $\boldsymbol{E}_a\,(x, y)$. We need to find out the radiated field in the half space $z > 0$.

This is a two-dimensional problem. We can employ the Fourier transform to obtain the radiation in the far field – this is normally a more efficient and effective method than the conventional current distribution approach, which uses the current distribution to characterize the antenna. Thus, the two-dimensional Fourier transform will be used to solve this problem.

It has been shown that the radiated field from such an aperture source is linked to a plane wave with vector amplitude $\boldsymbol{A}$ propagating in the direction of the propagation vector $\boldsymbol{k}$ by the following expression:

$$\boldsymbol{E}(x, y, z) = \frac{1}{4\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\boldsymbol{A}(k_x, k_y)e^{-j\boldsymbol{k}\cdot\boldsymbol{r}}dk_x dk_y \tag{5.48}$$

where

$$\boldsymbol{k}\cdot\boldsymbol{r} = (k_x\hat{\boldsymbol{x}} + k_y\hat{\boldsymbol{y}} + k_z\hat{\boldsymbol{z}})\cdot(\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}) = k_x x + k_y y + k_z z;$$

k_x, k_y, k_z are propagation constants along the x, y and z directions, respectively; they are linked to the wave number $\beta = \dfrac{2\pi}{\lambda} = \sqrt{{k_x}^2 + {k_y}^2 + {k_z}^2}$, and

$$k_x = \beta\sin\theta\cos\phi;\ k_y = \beta\sin\theta\sin\phi;\ k_z = \beta\cos\theta \tag{5.49}$$

$$\boldsymbol{A}(k_x, k_y) = \boldsymbol{A_t}(k_x, k_y) + A_z(k_x, k_y)\hat{\boldsymbol{z}};$$

$$A_z(k_x, k_y) = -\frac{\boldsymbol{k}\cdot\boldsymbol{A_t}(k_x, k_y)}{k_z} \text{ and}$$

$$\boldsymbol{A_t}(k_x, k_y) = A_x(k_x, k_y)\hat{\boldsymbol{x}} + A_y(k_x, k_y)\hat{\boldsymbol{y}} = \iint_{S_a}\boldsymbol{E_a}(x, y)e^{j(k_x x + k_y y)}dxdy \tag{5.50}$$

This is the Fourier transform of the aperture field. Thus, Equation (5.48) means that the radiated field $\boldsymbol{E}$ can be obtained using the Fourier transform of the aperture field. The calculation may be complicated since double integrals are involved.

In the far field, there is no field component along the propagation direction, i.e. TEM waves. Equation (5.48) can be much simplified to [1, 2]:

$$\boldsymbol{E}(\boldsymbol{r}) = j\beta \frac{e^{-j\beta r}}{2\pi r} \left[(A_x \cos\phi + A_y \sin\phi)\hat{\theta} + (A_y \cos\phi - A_x \sin\phi)\cos\theta \cdot \hat{\phi} \right] \quad (5.51)$$

A coordinate transform from Cartesian coordinates to spherical coordinates has been made here to simplify the expression of the field, which is inversely proportional to the distance, thus the radiated power is inversely proportional to the distance squared, which is the same as the radiated field from a short wire-type antenna. However, this radiated field has two equally important components in the θ and ϕ directions – this is not the case for most wire-type antennas.

The magnetic field in the far field (a local plane wave) is given by

$$\boldsymbol{H} = \hat{\boldsymbol{r}} \times \boldsymbol{E}/\eta \quad (5.52)$$

where η is the intrinsic impedance of the medium, as defined previously.

Thus, once the source aperture field distribution is known, the radiated fields can be obtained using these equations. Other relevant characteristics (such as the directivity and HPBW) can also be calculated using the formulas provided in Chapter 4. A summary of the characteristics of a few rectangular aperture antennas is presented in Figure 5.25, which shows clearly how

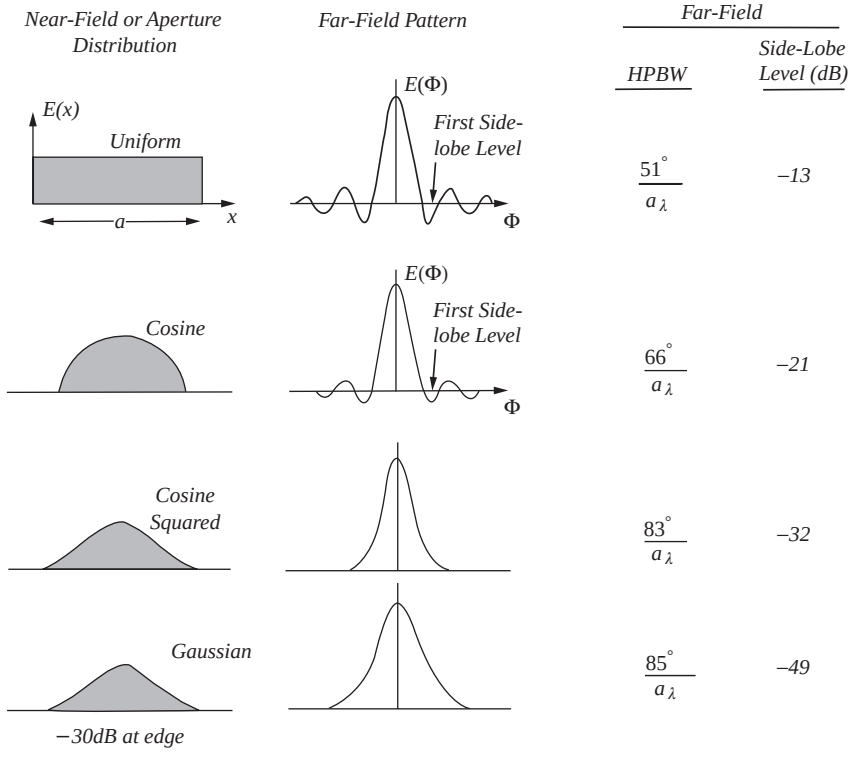


Figure 5.25 Aperture antenna radiation characteristics ($a_\lambda = a/\lambda$). (J. D. Kraus and R. J. Marhefka, Antennas for All Applications, 3rd edition, McGraw-Hill, Inc., 2002. Reproduced by permission of The McGraw-Hill Companies)

the aperture field distribution controls the radiation pattern. It is interesting to note that *there is a trade-off between the HPBW and the side-lobe level: the narrower the HPBW, the higher the side-lobe level*. It is also indicated that *the maximum aperture efficiency occurs for a uniform aperture distribution while the maximum beam efficiency (the ratio of the solid angle of the main beam to the total beam solid angle) occurs for a highly tapered aperture distribution* – a compromise is normally one of the most important design aspects.

It can be shown that, for aperture-type antennas, the directivity can also be obtained using the following formula, which is straightforward and powerful:

$$D = \frac{4\pi}{\lambda^2} \frac{|\iint_{S_a} \boldsymbol{E_a} ds|^2}{\iint_{S_a} |\boldsymbol{E_a}|^2 ds} \tag{5.53}$$

If the aperture distribution is of uniform amplitude, then this equation reduces to

$$D_u = \frac{4\pi}{\lambda^2} A_p \tag{5.54}$$

where A_p is the physical aperture area of the antenna, which was used to find the effective aperture size and efficiency in Chapter 4. Using Equation (4.22), we know that, in this case, the aperture efficiency is 100% and the directivity is the largest obtainable from such a physical size. This shows that *the larger the aperture, the larger the directivity – this is a general conclusion for aperture-type antennas.*

Example 5.4: Radiation from an open-ended waveguide. An open waveguide aperture of dimensions a along x and b along y located in the $z = 0$ plane is shown in Figure 2.32. The field in the aperture is TE_{10} mode and is given by

$$\boldsymbol{E_a} = \hat{\boldsymbol{y}} E_0 \cos\frac{\pi x}{a}, \quad |x| \le \frac{a}{2}, \text{ and } |y| \le \frac{b}{2}$$

a. find the radiated far field and plot the radiation pattern in both the E- and H-planes;
b. find the directivity.

Solution:

a. Using Equation (5.50) yields

$$\boldsymbol{A_t}(k_x, k_y) = \iint_{S_a} \boldsymbol{E_a}(x, y) e^{j(k_x x + k_y y)} dx dy = \hat{\boldsymbol{y}} E_0 \int_{-b/2}^{b/2} \int_{-a/2}^{a/2} \cos\frac{\pi x}{a} e^{j(k_x x + k_y y)} dx dy$$

Because

$$\int_{-b/2}^{b/2} e^{j(k_y y)} dy = \frac{1}{jk_y}(e^{jk_y b/2} - e^{-jk_y b/2}) = \frac{2}{k_y}\sin(k_y b/2) = b\,\mathrm{sinc}\,(k_y b/2)$$

and

$$\int_{-a/2}^{a/2} \cos\frac{\pi x}{a} e^{j(k_x x)} dx = 2\pi a \frac{\cos(k_x a/2)}{\pi^2 - (k_x a)^2},$$

We have

$$\boldsymbol{A_t}(k_x, k_y) = \hat{\boldsymbol{y}} 2\pi ab E_0 \sin c\,(k_y b/2) \frac{\cos(k_x a/2)}{\pi^2 - (k_x a)^2} \tag{5.55}$$

Note that the special function

$$\sin c\,(\alpha) = \frac{\sin\alpha}{\alpha} \tag{5.56}$$

Caution: some authors and Matlab (a well-known software package) define $\sin c\,(\alpha) = \dfrac{\sin\pi\alpha}{\pi\alpha}$, which is different from Equation (5.56).
The far field can be obtained using Equation (5.51) as:

$$\begin{aligned} \boldsymbol{E}(r) &= j\beta \frac{e^{-j\beta r}}{2\pi r} \left[A_y \sin\phi \cdot \hat{\theta} + A_y \cos\phi \cos\theta \cdot \hat{\phi} \right] \\ &= j\beta \frac{e^{-j\beta r}}{r} ab E_0 \sin c\,(k_y b/2) \frac{\cos(k_x a/2)}{\pi^2 - (k_x a)^2} \left[\sin\phi \cdot \hat{\theta} + \cos\phi \cos\theta \cdot \hat{\phi} \right] \end{aligned} \tag{5.57}$$

In the $\phi = 0$ plane (H-plane), i.e. the *xz* plane, we have $k_x = \beta \sin\theta$, $k_y = 0$, and

$$\begin{aligned} \boldsymbol{E}(r) &= j\beta \frac{e^{-j\beta r}}{r} ab E_0 \frac{\cos(k_x a/2)}{\pi^2 - (k_x a)^2} \left[\cos\theta \cdot \hat{\phi} \right] \\ &= j\beta ab E_0 \frac{e^{-j\beta r}}{r} \frac{\cos(\beta a \sin\theta/2)}{\pi^2 - (\beta a \sin\theta)^2} \cos\theta \cdot \hat{\phi} \end{aligned} \tag{5.58}$$

In this plane, the radiated field is actually $\boldsymbol{E}_y(r)$, parallel to the aperture field $\boldsymbol{E}_a$.
While in the $\phi = \pi/2$ plane (the E-plane), i.e. the *yz* plane, we have $k_x = 0$, $k_y = \beta \sin\theta$, and

$$\boldsymbol{E}(r) = j\beta \frac{e^{-j\beta r}}{r} ab E_0 \sin c\,(k_y b/2) \frac{1}{\pi^2} \left[\hat{\theta} \right] = j\beta ab E_0 \frac{e^{-j\beta r}}{\pi^2 r} \sin c \left(\frac{\beta b \sin\theta}{2} \right) \cdot \hat{\theta} \tag{5.59}$$

The first side-lobe level for the *sinc* function is −13.2 dB, which is the value for a uniform aperture distribution (along the *y* direction).
For a standard waveguide, as given in Table 2.5, the typical values of the central frequency for βa and βb are:

$$\beta a \approx \frac{4\pi}{3};\ \text{and}\ \beta b \approx \frac{2\pi}{3}$$

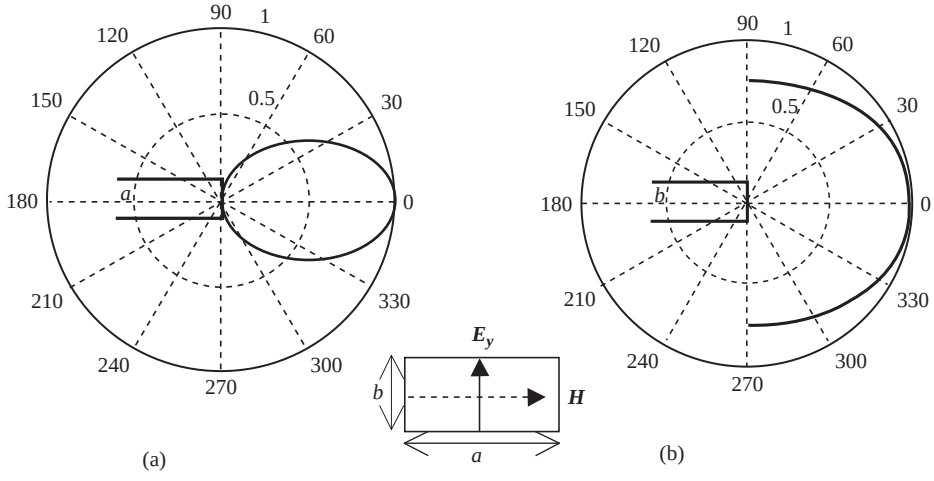


Figure 5.26 Typical radiation patterns of an open waveguide in (a) the H-plane; (b) the E-plane

Thus, the far-field radiation patterns at the two principal planes can be plotted as in Figure 5.26. As expected, it has a unidirectional pattern and the maximum radiation is normal to the waveguide open end. The pattern in the H-plane is close to $\cos\theta$, while the pattern in the E-plane is very broad. It is also clearly shown from the E-plane pattern that a ground plane is required, otherwise the field will be radiated to the back of the open waveguide.

b. The directivity can be obtained using Equation (5.53), i.e.

$$D = \frac{4\pi}{\lambda^2}\frac{\left|\iint_{S_a} \boldsymbol{E_a} ds\right|^2}{\iint_{S_a} |\boldsymbol{E_a}|^2 ds} = \frac{4\pi}{\lambda^2}\frac{\left|\int_{-b/2}^{b/2}\int_{-a/2}^{a/2} E_0 \cos\frac{\pi x}{a} dxdy\right|^2}{\int_{-b/2}^{b/2}\int_{-a/2}^{a/2}\left|E_0 \cos\frac{\pi x}{a}\right|^2 dxdy} = \frac{4\pi}{\lambda^2}\frac{8}{\pi^2}ab \approx \frac{4\pi}{\lambda^2}(0.81ab)$$

Again it is proportional to the aperture's physical size (ab) and this result indicates that the aperture efficiency is reduced to 81% compared with uniform excitation. Since the waveguide size is small, the directivity is small. The only way to increase it is to flare out its ends into a horn. Figure 5.27 shows three different horns with increased directivity.

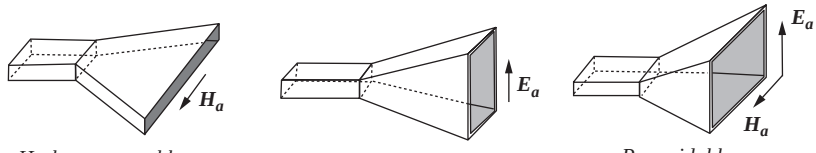


Figure 5.27 Three horn antennas

5.2.2 Horn Antennas

There is no doubt that horn antennas are the simplest and one of the most widely used forms of microwave antenna – the antenna is nicely integrated with the feed line (waveguide) and the performance can be controlled easily. They are mainly used for standard antenna gain and field measurements, feed elements for reflector antennas and microwave communications. The horn can take many different forms: pyramidal horns (shown in Figure 5.27) and conical horns are the most popular types – the former is most suitable for linear polarization and the latter for circular polarization. Since the basic theory is covered in the previous section, the focus of this section is on design and performance estimation.

5.2.2.1 Pyramidal Horns

The open-ended waveguide has a small directivity and broad beamwidth, as demonstrated in Example 5.4. Thus, it is not suitable for most practical applications. The pyramidal horn has therefore evolved from the open waveguide and it is flared in both the E- and H-planes, as shown in Figure 5.28, which results in narrow beamwidths in both principal planes. The question is how to obtain the optimum design for a specified gain/directivity at a desired operating frequency.

Ideally, the phase of the field across the horn mouth should be constant in order to obtain the desired pattern with minimized side lobes. This requires a very long horn. However, the horn should be as short as possible for practical convenience. *An optimum design is therefore a compromise in which the difference in the path length along the edge, l_E, and the center of the horn, R_2, is made about 0.25λ.*

All dimensional parameters are shown in Figure 5.28. The directivity to be achieved is D at the operational wavelength λ. The feed waveguide dimensions are of width a and height b.

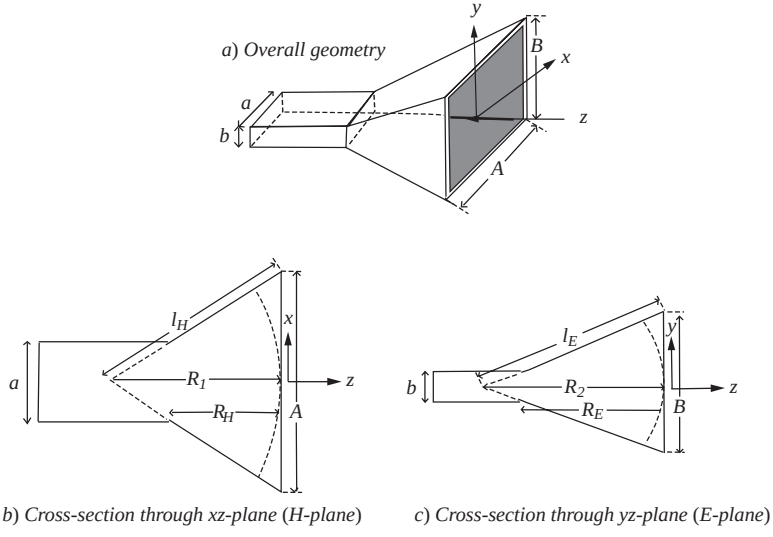


Figure 5.28 Pyramidal horn antennas with dimensional parameters: (a) Overall geometry (b) Cross-section through *xz*-plane (H-plane) (c) Cross-section through *yz*-plane (E-plane)

To make such a pyramidal horn, we need to know three of the parameters (A, l_H, R_1 and R_H) in the H-plane and three of the parameters (B, l_E, R_2 and R_E) in the E-plane. The objective of the design is to determine these unknown dimensions.

In the H-plane, the dimensions are linked by

$$\begin{aligned} l_H^2 &= R_1^2 + (A/2)^2 \\ R_H &= (A-a)\sqrt{(l_H/A)^2 - 0.25} \end{aligned} \tag{5.60}$$

and the maximum phase difference in the H-plane is

$$l_H - R_1 \approx \frac{1}{2R_1}\left(\frac{A}{2}\right)^2 = t\lambda \tag{5.61}$$

where t is the maximum phase deviation over the wavelength in the H-plane, i.e. $t = (l_H - R_1)/\lambda$. The larger the value of t, the broader the beamwidth. It has been found that $t = 3/8 = 0.375$ (but 0.25 for the E-plane) gives the optimum design [2, 3, 12, 20], which means

$$A = \sqrt{3\lambda R_1} \tag{5.62}$$

In the E-plane, the dimensions are linked by

$$\begin{aligned} l_E^2 &= R_2^2 + (B/2)^2 \\ R_E &= (B-b)\sqrt{(l_E/B)^2 - 0.25} \end{aligned} \tag{5.63}$$

and the maximum phase difference in the E-plane is

$$l_E - R_2 \approx \frac{1}{2R_2}\left(\frac{B}{2}\right)^2 = s\lambda \tag{5.64}$$

where s is the maximum phase deviation in the E-plane; the larger the value of s, the broader the beamwidth. It has been found that $s = 1/4 = 0.25$ gives the optimum design, which means

$$B = \sqrt{2\lambda R_2} \tag{5.65}$$

When both the E- and H-planes are put together to form the pyramidal horn, the following condition has to be satisfied in order to make it physically realizable and properly connected to the feed waveguide:

$$R_E = R_H \tag{5.66}$$

The directivity is related to the aperture efficiency factor η_{ap} and aperture size AB by Equation (4.22), that is

$$D = \frac{4\pi}{\lambda^2}(\eta_{ap} AB) \tag{5.67}$$

For the optimum gain horn, $\eta_{ap} \approx 0.51 = 51\%$.

We have eight independent equations (5.60a, 5.60b, 5.61, 5.63a, 5.63b, 5.64, 5.66 and 5.67) to solve eight unknown variables in principle for the design problem to be resolved. Since most of them are not linear equations, it is actually difficult (if not impossible) to obtain a solution analytically. After some mathematical manipulation, we can yield

$$(B-b)\frac{B}{s} = (A-a)\frac{A}{t} \tag{5.68}$$

The solution to B (which is positive, thus the other solution is not useful) is

$$B = \frac{b + \sqrt{b^2 + 4sA(A-a)/t}}{2} \tag{5.69}$$

Replacing B in Equation (5.67) gives a desired design equation:

$$A^4 - aA^3 + \frac{tbD\lambda^2}{4\pi s\eta_{ap}}A = \frac{tD^2\lambda^4}{16s\pi^2\eta_{ap}^2} \tag{5.70}$$

This fourth-order equation in A can be solved using numerical methods. Alternatively, it may also be attempted by trial and error. For the optimum design, use a first guess approximation [12, 20]:

$$A = 0.45\lambda\sqrt{D} \tag{5.71}$$

Once the value of A is obtained, we can easily calculate other dimensions of the structure. The half-power beamwidth for an optimum horn in the H-plane is [12]

$$HP_H = \phi_{HP} \approx 78^\circ\frac{\lambda}{A} \tag{5.72}$$

In the E-plane it is

$$HP_E = \theta_{HP} \approx 54^\circ\frac{\lambda}{B} \tag{5.73}$$

These are approximate values; slightly different approximations are given in [4]. It is also noted that the directivity obtained using these equations and Equation (4.14) is different from that using Equation (5.67) unless $\eta_{ap} = 0.78$ (it should be 0.51 for the optimized horn). This may be due to the fact that Equation (4.14) is an approximation for a single-lobe antenna and this one has side lobes.

Now we are going to use the following example to summarize the procedure of designing an optimum pyramidal horn.

Example 5.5: Optimum horn design. Design a standard gain horn with a directivity of 20 dBi at 10 GHz. A WR-90 waveguide will be used to feed the horn.

Solution:
The directivity is $D = 20$ dBi $= 100$, wavelength $\lambda = 30$ mm and the dimensions of the waveguide are $a = 22.86$ mm and $b = 10.16$ mm.

Step 1: *Compute the dimension A from the design Equation (5.70).*
As suggested above, the parameters for the optimum horns are:

$$\eta_{ap} = 0.51;\ s = 0.25;\ t = 0.375$$

and the design Equation (5.70) becomes

$$A^4 - 22.86A^3 + 214020A = 2.9581 \times 10^8$$

Using Equation (5.71), we obtain an initial guessing value:

$$A = 0.45\lambda\sqrt{D} = 135\ (\text{mm})$$

and plot the function

$$Y(A) = A^4 - 22.86A^3 + 21402A - 2.9581 \times 10^8$$

around $A = 135$ (mm). As seen in Figure 5.29, the actual solution is readily found as

$$A \approx 133.96\,(\text{mm})$$

Step 2: *Use Equation* (5.69) *to yield B*
In this case $B = 104.82$ mm

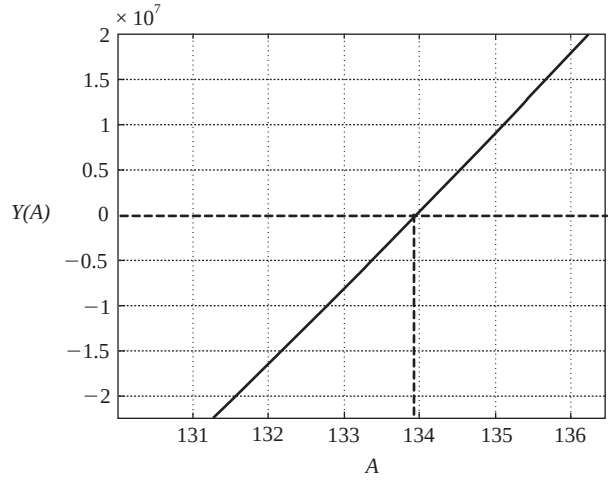


Figure 5.29 Function Y(A) vs. *A*

Step 3: *Find the remaining dimensions from:*
Equation (5.61) *or* (5.62) *for* R_1;
Equation (5.64) *or* (5.65) *for* R_2;
Equation (5.60) *for* l_H *and* R_H;
Equation (5.63) *for* l_E *and* R_E.
We get
$R_1 = 199.41$mm;
$R_2 =$183.14 mm;
$l_H = 210.36$ mm and $R_H = 165.38$ mm;
$l_E = 190.49$ mm and $R_E = 165.38$ mm.
Step 4: *Check if* R_H *and* R_E *are the same. If not, it means that the solution of A in Step 1 is not accurate enough.*

From the results in Step 3, we can see that R_H and R_E are identical, thus the design is very good. However, if we used the guessing value $A = 0.45\lambda\sqrt{D} = 135$ (mm) as the solution, it would give $R_H = 168.21$ mm and $R_E = 162.73$ mm. They are obviously different, which means that the design needs to be revised.

As the TE_{10} mode is the field pattern for such an antenna, the aperture field distribution is still the same as that of an open-ended waveguide (TE_{10} mode as well) but with a different phase term, this is

$$\boldsymbol{E_a} = \hat{\boldsymbol{y}} E_0 \cos\frac{\pi x}{A} e^{-j\beta/2(x^2/R_1 + y^2/R_2)}, \quad |x| \le \frac{A}{2}, \text{ and } |y| \le \frac{B}{2} \tag{5.74}$$

The radiation pattern is closely related to the phase error parameters s and t. There is no simple expression for the pattern. Universal radiation patterns, as shown in Figure 5.30 where a weighting factor $(1 + \cos\theta)/2$ is not included, have been produced for design purposes and a detailed treatment of this subject can be found in [2]. The radiation pattern of this design is shown in Figure 5.31. The side lobes in the E-plane are much higher than those in the H-plane – a common feature of this type of antenna, since the aperture field in the E-plane is uniform (not tapered).

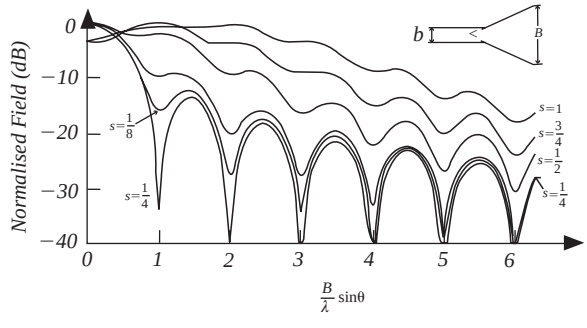


Figure 5.30 E-plane universal patterns for E-plane sectorial and pyramidal horns

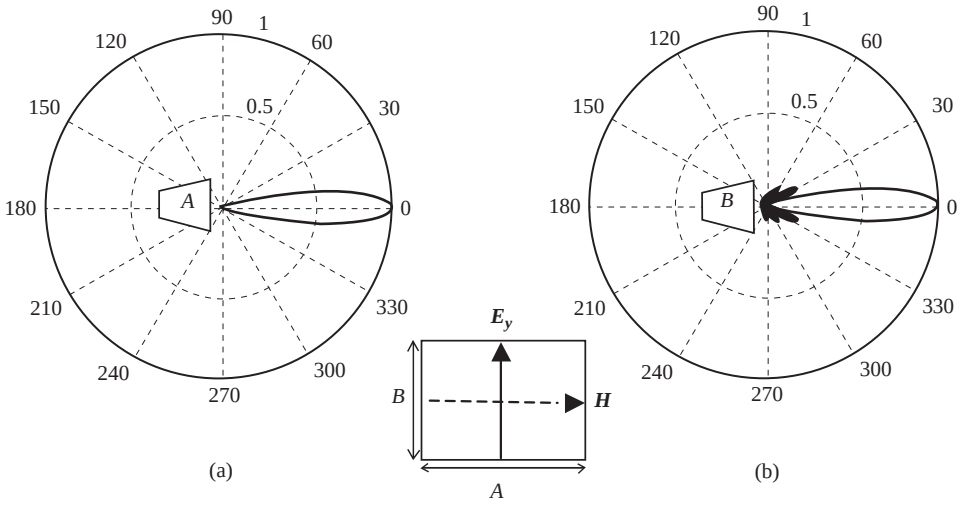


Figure 5.31 Radiation patterns of a pyramidal horn antenna (a) H-plane; (b) E-plane

The half-power beamwidths in the two principal planes for this antenna are

$$HPBW_H = \phi_{HP} \approx 78^\circ \frac{\lambda}{A} = 17.47^\circ$$

$$HPBW_E = \theta_{HP} \approx 54^\circ \frac{\lambda}{B} = 15.45^\circ$$

It should be pointed out that:

- The optimum horn is designed for a specific frequency and gain.
- For a horn of fixed length, there is a maximum directivity/gain obtainable, which normally corresponds to the optimum design. Under such a condition, the directivity is proportional to the aperture size. Otherwise, this conclusion may not be valid. For example, a horn with a larger aperture size than the optimal one has larger phase error parameters, which result in smaller directivity.
- There are two special cases: in Figure 5.28, if $A = a$, the structure is called the *E-plane sectorial horn antenna*; if $B = b$, the structure is called the *H-plane sectorial horn antenna*. They are both shown in Figure 5.27. They have a very broad beamwidth in one of the principal planes and a very narrow beamwidth in the other plane (the plane opened up). The design procedure is the same as that for the pyramidal horn.

5.2.2.2 Other Horn Antennas

In addition to the pyramidal horn, there are many other horn antennas. The well-known ones include:

- The *circular horn*: a simple circularly polarized antenna.

- The *corrugated horn*: this can provide reduced edge diffraction and cross polarization, together with improved aperture efficiency (to 75–80%) and radiation pattern symmetry.
- The *single- or double-ridged horn*: a directional broadband antenna using a tapered ridge, as shown in Figure 5.23.

Detailed discussion of these antennas can be found in references such as [3].

5.2.3 Reflector and Lens Antennas

The typical gain of a practical horn antenna is up to 20 dBi or so. Higher gains are obtainable, but it means a much larger and heavier horn, which is not suitable for most applications. As an alternative in practice, reflector and lens antennas can offer much higher gains than horn antennas and are normally relatively easy to design and construct. They are probably the most widely used antennas for high-frequency and high-gain applications in radio astronomy, radar, microwave and millimeter wave communications and satellite tracking and communications.

Although reflector and lens antennas can take various geometrical configurations, the most popular shape is the paraboloid – because of its excellent ability to produce a pencil beam (high gain) with low side lobes and good cross-polarization characteristics in the radiation pattern. The largest fully steerable reflector in the world is the 100 m diameter radio telescope of the Max Planck Institute for Radioastronomy at Effelsberg in Germany, whereas the largest reflector antenna is the 305 m radio telescope at Arecibo Observatory in the USA at the time of writing. As displayed in Figure 5.32, the typical feed for such an antenna is a horn antenna, which can be placed at the focal point in front of the reflector (front-fed) or at the back (vertex) of the reflector (Cassegrain-fed). The latter is known as the *Cassegrain reflector antenna* and was invented to avoid using a long feed line and to minimize the feed blockage problems of a conventional paraboloidal antenna (the side lobes can be reduced as well). Due to its popularity, the emphasis of this section will be on the paraboloidal reflector antenna. The design principles can be applied to other reflector antenna designs. Many research and design papers have been published in this area and some of the most referenced papers can be found in [21].

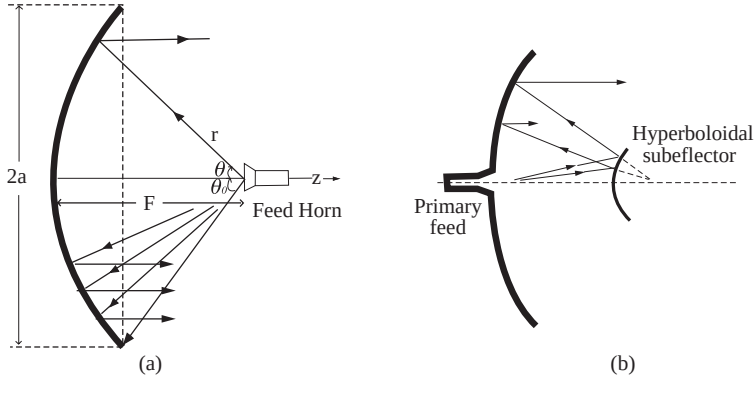


Figure 5.32 (a) Paraboloidal and (b) Cassegrain reflector antennas (R. E. Collin, Antennas and Radiowave Propagation. McGraw-Hill, Inc., 1985. Reproduced by permission of The McGraw-Hill Companies)

5.2.3.1 Paraboloidal Reflector Characteristics

As Figure 5.32 shows, a front-fed paraboloidal antenna consists of a reflector and a feed antenna. The surface of a paraboloidal reflector, which is illuminated by the feed antenna, is formed by rotating a parabola about its axis. Its surface is therefore a paraboloid of revolution and rays emanating from the focus of the reflector are transformed into plane waves – this means that it is highly directional. If the reflector was extremely large (infinite) and the feed was a point source at the focal point having radiation only towards the paraboloid, all the radiated EM energy would be directed in one direction (the z-direction in Figure 5.32) with zero beamwidth. Here we have assumed that the radiated waves can be treated as rays (otherwise the conclusion would not be valid), i.e. the structure is much greater than the wavelength. This unique feature of the paraboloid has made it extremely suitable for high-frequency and high-gain applications.

5.2.3.2 Analysis and Design

In practice, it is not possible to make the reflector infinitely large (actually we always try to make it as small as possible) and truncation has to take place. Also, the feed antenna cannot be a point source, which means that the actual performance of the antenna will be different from the ideal one. Designing a good paraboloidal antenna is, in fact, a difficult task.

Let us assume that the diameter of the reflector is $2a$ and the focal length is F. Any point on this paraboloid must satisfy the following condition:

$$r = \frac{2F}{1+\cos\theta} = F\sec^2(\theta/2) \qquad \theta \le \theta_0 \tag{5.75}$$

where the subtended/angular aperture angle θ_0 (also known as the *edge angle*) is determined by the reflector diameter and the focal length:

$$\theta_0 = \tan^{-1}\left|\frac{a}{F - a^2/(4F)}\right| \tag{5.76}$$

Here there are two parameters, diameter $2a$ and focal length F, to be decided, which does not seem to be too bad. However, another element of the reflector antenna is the feed, which is an antenna on its own and has many variables. Thus, the complete design of the reflector antenna is actually a complex task since there are many parameters which could be changed to meet (or fail) the specifications, some are independent and some are interlinked. *The reflector design problem consists primarily of matching the feed antenna pattern to the reflector. The usual goal is to have the feed pattern about 10 dB down in the direction of the rim*, that is the *edge taper* = (the field at the edge)/(the field at the center) $\approx$ 10 dB. Feed antennas with this property can be constructed for the commonly used $F/2a$ value of 0.3 to 1.0. Higher values lead to better cross-polarization performance, but require a narrower feed pattern and hence a physically larger feed antenna.

The antenna analysis can be conducted using the Fourier-transform based *aperture distribution method* introduced in Section 5.1. Alternatively, the *current distribution method* commonly used for wire-type antenna analysis and the equivalence principle can be employed to obtain the equivalent source current at the aperture. Both methods yield results that are essentially the same for the principal polarized radiation pattern and agree quite well with experimental results as long as the aperture is large enough in terms of the wavelength. Detailed analysis can

be found in [21]. As a high-gain antenna, the most important parameter is obviously the directivity/gain. Thus, we are going to see how to estimate this parameter without using expensive computer simulation tools.

Aperture Efficiency and Directivity

Since the feed antenna is closely linked to the reflector, the aperture efficiency should surely reflect this linkage. Let $g(\theta)$ be the power radiation pattern of the feed located at the focus – it is circularly symmetrical (not a function of ϕ). It has been shown that the aperture efficiency is given as [2, 21]

$$\eta_{ap} = \cot^2\left(\frac{\theta_0}{2}\right)\left|\int_0^{\theta_0} g(\theta)\tan\left(\frac{\theta}{2}\right)d\theta\right|^2 \tag{5.77}$$

As expected, this is determined by both the reflector and the feed. The maximum aperture efficiency is around 82%, which is higher than that of a pyramidal horn.

The aperture efficiency is generally viewed as the product of the

1. **Spillover efficiency:** the fraction of the total power intercepted and collimated by the reflector; it reduces the gain and increases the side-lobe levels.
2. **Taper efficiency:** the uniformity of the amplitude distribution of the feed pattern over the surface of the reflector.
3. **Phase efficiency:** the phase uniformity of the field over the aperture plane; it affects the gain and side lobes.
4. **Polarization efficiency:** the polarization uniformity of the field over the aperture plane.
5. **Blockage efficiency:** by the feed; it reduces gain and increases side-lobe levels. The support structure can also contribute to the blockage.
6. **Random error efficiency:** over the reflector surface.

Some of these requirements contradict each other. For example, the larger the reflector angular aperture angle θ_0, the larger the spillover efficiency, but the smaller the taper efficiency. The optimum aperture efficiency is the best trade-off from all aspects.

Once the aperture efficiency is found, the directivity can be readily obtained using Equation (4.22):

$$D = \frac{4\pi}{\lambda^2}\eta_{ap}(\pi a^2) \tag{5.78}$$

The analysis here has not taken the feed antenna efficiency into account, which is about 70–80% if it is a horn antenna. Thus, the overall reflector efficiency factor is in the region of 50–70%. It is very common to make a reflector antenna with a gain of over 30 dBi. Some of the world's largest antennas have a gain over 70 or even 80 dBi, which is almost impossible for a single wire-type antenna to achieve – remember the typical gain for a Yagi–Uda or log-periodic antenna is about 10–15 dBi.

The HPBW can be estimated by the rule of thumb:

$$HPBW \approx 70^\circ\frac{\lambda}{2a} \tag{5.79}$$

The beamwidth also depends on the edge illumination. Typically, as the edge attenuation increases, the beamwidth widens and the side lobes decrease.

Design Considerations and Procedures

In addition to the aperture efficiency and directivity/gain, cross-polarization, phase errors (the maximum fractional reduction in directivity is $\delta^2(1-\delta^2/4)$, where δ is the phase error), HPBW (which can be estimated using Equation (4.14) once the directivity is known) and side lobes must also be considered in design. The priority is really down to the specific application.

The design procedure starts with selection of the feed antenna, which determines the antenna polarization and the reflector F and $2a$. As mentioned earlier, the usual goal is to let the feed pattern be about 10 dB down in the direction of the reflector rim. The feed and reflector are interlinked; an iterative process may be required to ensure that the feed antenna pattern is well matched with the reflector. Once the feed pattern and the reflector are known, the radiated field can be calculated, where complex integrations are involved [2, 21]. Normally, computer software is required to accomplish such a task. For example, the software package GRASP, a product of TICRA, is well suited for this application. This package is a set of tools for analyzing reflector antennas and antenna farms. The program is based on PO (physical optics)/GTD (geometrical theory of diffraction) calculation. More about software for antenna designs will be given in Chapter 6.

Example 5.6: Edge taper and spillover efficiency. A circular parabolic reflector has $F/2a = 0.5$. The field pattern of the feed antenna is $E(\theta) = \cos\theta$, $\theta < \pi/2$. Find the edge taper, spillover efficiency and aperture efficiency.

Solution:

The edge angle is given by Equation (5.76):

$$\theta_0 = \tan^{-1}\left|\frac{a}{F - a^2/(4F)}\right| = \tan^{-1}\left|\frac{(F/2a)/2}{(F/2a)^2 - 1/16}\right| = \tan^{-1}\left|\frac{0.25}{(0.5)^2 - 1/16}\right| = 53.13^\circ$$

The aperture illumination function is

$$A(\theta) = E(\theta)\frac{1}{r}e^{-j\beta r} = \cos(\theta)\frac{1}{r}e^{-j\beta r}$$

Since r is given by Equation (5.75),

$$|A(\theta)| = \frac{\cos(\theta)(1+\cos\theta)}{2F}$$

The edge taper is therefore:

$$\frac{A(\theta_0)}{A(0)} = \frac{\cos(\theta_0)(1+\cos(\theta_0))}{(1+1)} = 0.4800 = -6.3752\,\text{dB}$$

The spillover efficiency is

$$\eta_s = \frac{P_{intercepted}}{P_{radiated}} = \frac{\int_0^{\theta_0} g(\theta)\sin\theta d\theta}{\int_0^{\pi} g(\theta)\sin\theta d\theta} = \frac{\int_0^{\theta_0} E(\theta)^2\sin\theta d\theta}{\int_0^{\pi} E(\theta)^2\sin\theta d\theta} = \frac{\int_0^{\theta_0} \cos^2(\theta)\sin\theta d\theta}{\int_0^{\pi} \cos^2(\theta)\sin\theta d\theta} \approx 0.78$$

Thus, the spillover loss is about -1.08 dB.

The aperture efficiency can be calculated using Equation (5.77). It is about 0.59.

If a uniform aperture illumination is needed, we have:

$$\frac{A(\theta)}{A(0)} = \frac{E(\theta)(1+\mathbf{cos}(\theta))}{E(0)(1+1)} \equiv 1$$

The feed antenna pattern should be

$$E(\theta) = \frac{2}{1+\mathbf{cos}\theta} \tag{5.80}$$

This means a trough at the center ($\theta = 0$), which is hard to achieve in practice.

5.2.3.3 Offset Parabolic Reflectors

To eliminate some of the deficiencies of symmetrical configurations, the offset parabolic reflector design has been developed for single- and dual-reflector systems, which reduces aperture blockage and offers the advantage of allowing the use of a larger $F/(2a)$ ratio while maintaining acceptable structure rigidity. The configuration is also widely employed, most noticeably for domestic satellite receiving antennas, which have a typical gain of about 30 dBi. It is also utilized in radar and other applications.

Figure 5.33 is an example of an offset parabolic reflector used for a radar positioning application where the HPBW is required to be very narrow in the horizontal plane but wider in

Figure 5.33 An example of an offset parabolic reflector radar antenna (Reproduced by permission of Guidance Ltd, UK)

the vertical plane. GRASP was employed to predict the performance. The simulated radiation patterns are shown in Figure 5.34 and the main specifications for the H-plane are:

- the gain: 28.0 dB (27.9 dB, measured)
- the half-power beamwidth: 4.0 degrees (4.4 degrees)
- the 10 dB beamwidth: 6.8 degrees (7.6 degrees)
- the 1st side-lobe level: −19.5 dB (−26.3 dB)

Measured results are given in the brackets. It is apparent that the simulated results are very close to the measured ones (except for the side lobes in the H-plane). This indicates that our simulation and design tool worked well for this antenna.

Some other reflector antennas, such as flat sheet reflector antennas and corner reflectors, are also used in practice.

5.2.3.4 Lens Antennas

Just like reflector antennas, lens antennas can also convert a spherical wave into a plane wave to produce high gains and pencil beams. They are only suitable for high-frequency (> 4 GHz) applications. There are basically two types, as shown in 5.35: delay lenses, in which the electrical path length is increased by the lens medium (using low-loss dielectrics with a relative permittivity greater than one, such as Lucite or polystyrene), and fast lenses, in which the electrical path length is decreased by the lens medium (using metallic or artificial dielectrics with a relative permittivity smaller than one). The source or primary antenna is normally a tapered horn antenna as for a reflector antenna. Lens antennas have very similar characteristics to the paraboloidal reflector antennas discussed earlier in this section. But they do offer several advantages over reflector antennas, which make them more attractive for some aerospace applications such as Earth observation and radars. These advantages include wide-angle scan, very low side lobes and low feed horn blockage. A very good discussion on lens antennas can be found in [4].

5.2.4 Slot Antennas and Babinet's Principle

Slot antennas can be considered a very special group of aperture-type antennas. They are very low-profile and can be conformed to basically any configuration, thus they have found many applications, for example on aircraft and missiles. They may be conveniently energized with a coaxial transmission line or a waveguide, as Figure 5.36 shows. The open-end waveguide analyzed in Example 5.4 may be viewed as a slot antenna if a ground plane is used. The aperture distribution method can be employed again to yield the radiated fields. For narrow slots, there is another way to do it: we are going to use the equivalence principle to obtain the radiation characteristic. This is simple and convenient and the knowledge gained for wire-type antennas can be readily applied to this case.

Recall the equivalence principle introduced in Chapter 3, the radiated field by the slot is the same as the field radiated by its equivalent surface electric current and magnetic current, which

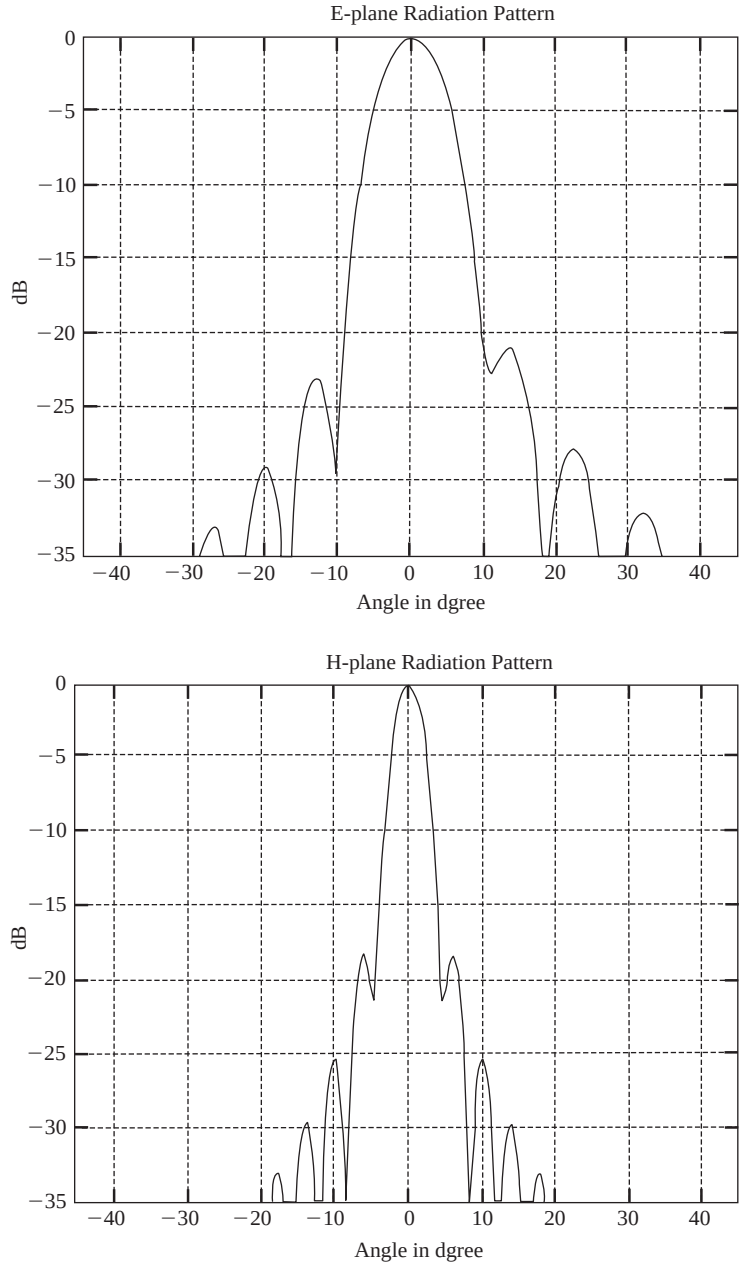


Figure 5.34 Computed radiation patterns in the E- and H-planes

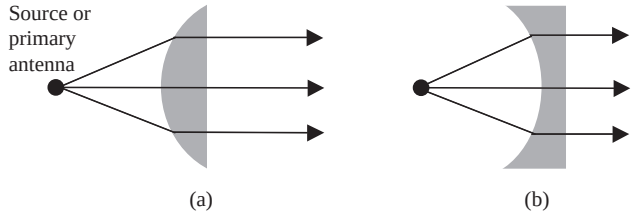


Figure 5.35 Two types of lens antenna (a) delay lens with relative permittivity $\varepsilon_r > 1$; (b) fast lens with relative permittivity $\varepsilon_r < 1$

were given by Equation (3.37), that is

$$\boldsymbol{J_S} = \hat{\boldsymbol{n}} \times \boldsymbol{H},\ \boldsymbol{M_S} = -\hat{\boldsymbol{n}} \times \mathbf{E}$$

where $\boldsymbol{E}$ and $\boldsymbol{H}$ are the electric and magnetic fields within the slot and $\hat{\boldsymbol{n}}$ is the unit vector normal to the slot surface S.

Take a half-wavelength slot as an example. Since its equivalent electric surface current $\boldsymbol{J_S} = \hat{\boldsymbol{n}} \times \boldsymbol{H} = 0$, the remaining source at the slot is its equivalent magnetic current $\boldsymbol{M_S} = -\hat{\boldsymbol{n}} \times \boldsymbol{E}$ (it would be $2\boldsymbol{M_S}$ if the conducting ground plane were removed using the imaging theory). This is the same kind of equivalent magnetic source as a loop, as discussed in Section 5.1.3. The current distribution along this half-wavelength magnetic source is the same as the current distribution along a half-wavelength dipole. We can therefore use the duality principle introduced in Section 5.1.3 to find the radiated field of the slot from the dipole field.

The normalized electric and magnetic fields of a half-wavelength dipole can be obtained using Equations (5.3) and (5.4) as

$$E_\theta = \frac{\cos[\pi/2\cos\theta]}{\sin\theta},\ \text{and}\ H_\phi = \frac{\cos[\pi/2\cos\theta]}{\sin\theta} \tag{5.81}$$

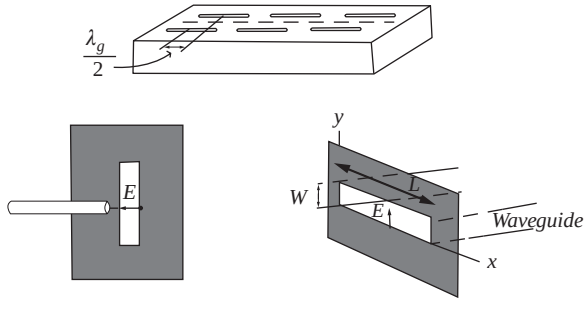


Figure 5.36 Slot antennas fed by a coaxial cable and a waveguide

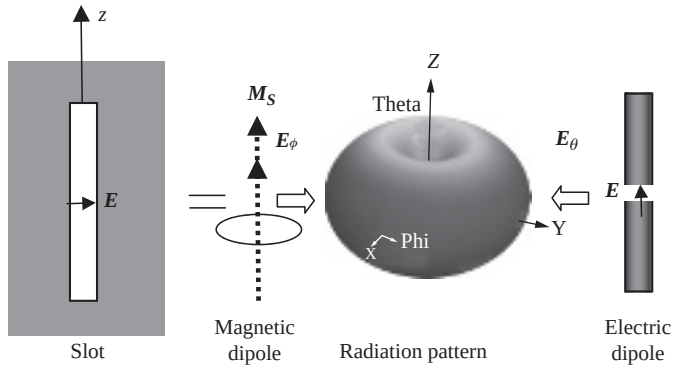


Figure 5.37 A slot antenna, its radiation pattern and its complementary dipole

Using the duality relation in Table 5.3, we obtain the normalized radiated field from a half-wavelength slot antenna as

$$E_\phi = \frac{\cos[\pi/2\cos\theta]}{\sin\theta}, \text{ and } H_\theta = \frac{\cos[\pi/2\cos\theta]}{\sin\theta} \tag{5.82}$$

Thus, the only change in the field from dipole to slot is that E_θ is replaced by E_ϕ, and H_ϕ is replaced by H_θ. The radiation pattern shape is not changed, as seen in Figure 5.37. Of course, if the slot antenna only radiates to half of the space, then the radiation pattern should only appear in half of the space.

Another very important parameter of the antenna is the input impedance. Again, if we utilize another theory, Babinet's principle, this parameter can also be obtained easily using our existing knowledge. *Babinet's principle* states that *the field at any point behind a plane having a screen, if added to the field at the same point when the complementary screen is substituted, is equal to the field at the point when no screen is present.* Obviously it is meant for optics. If this principle is extended to antennas and radio propagation and the equivalent transmission line model is employed, as we did in Chapter 3 for radio wave transmission and reflection analysis, we can get the following important relationship for the input impedances of complementary antennas:

$$Z_{slot}Z_{dipole} = \frac{\eta^2}{4} \tag{5.83}$$

This means that the product of the impedances of two complementary antennas is the constant $\eta^2/4$ and this is the interpretation of Babinet's principle in antennas.

Since the impedance for a half-wavelength dipole is about 73 ohms, the corresponding slot has an impedance of

$$Z_{slot} = \frac{\eta^2}{4Z_{dipole}} = \frac{377}{4\times 73} \approx 486\,(\Omega)$$

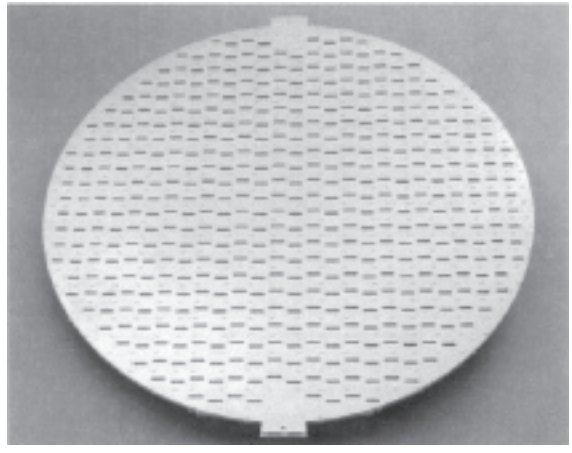


Figure 5.38 An airborne slot waveguide antenna array

We discussed frequency-independent antennas earlier in this chapter, and the self-complementary configurations shown in Figure 5.22 are two of them. Because their impedances are the same, in free space this type of antenna has a constant impedance of

$$Z_a = \sqrt{\frac{\eta^2}{4}} = 188.5\,(\Omega) \tag{5.84}$$

Slot antennas, especially slot waveguide antennas where the antenna and feed line are integrated and exhibit high power-handling capacity, are often used to form an antenna array for applications such as airborne radar. Figure 5.38 is such an example: 26 waveguides and over 300 slot elements are employed to form this antenna array. Each slot is carefully designed to ensure that the field aperture distribution follows a Gaussian distribution, as shown in Figure 5.25. It has a very high gain (over 35 dBi) and low side lobes (below −25 dB). More will be presented on antenna arrays in Section 5.3.

5.2.5 Microstrip Antennas

A microstrip antenna, also known as a patch antenna, consists of a metal patch on a substrate on a ground plane, as shown in Figure 5.39. Different feed configurations, including aperture-coupled, microstrip line feed and coaxial feed, are also shown in the figure. The patch can take various forms to meet different design requirements. Typical shapes are rectangular, square, circular and circular ring. The microstrip antenna is low-profile, conformable to planar and nonplanar surfaces, simple and cheap to manufacture using modern printed-circuit technology, compatible with MMIC (monolithic microwave integrated circuit) designs and mechanically robust. In addition, it is very versatile in terms of resonant frequency, input impedance, radiation pattern and polarization. All these have made it an extremely popular modern antenna for frequencies above 300 MHz (from the UHF band). A huge number of research papers and many books have been published over the years in this area [22–25]. The major disadvantages of this type of antenna are: low efficiency (conducting, dielectric and especially surface wave losses),

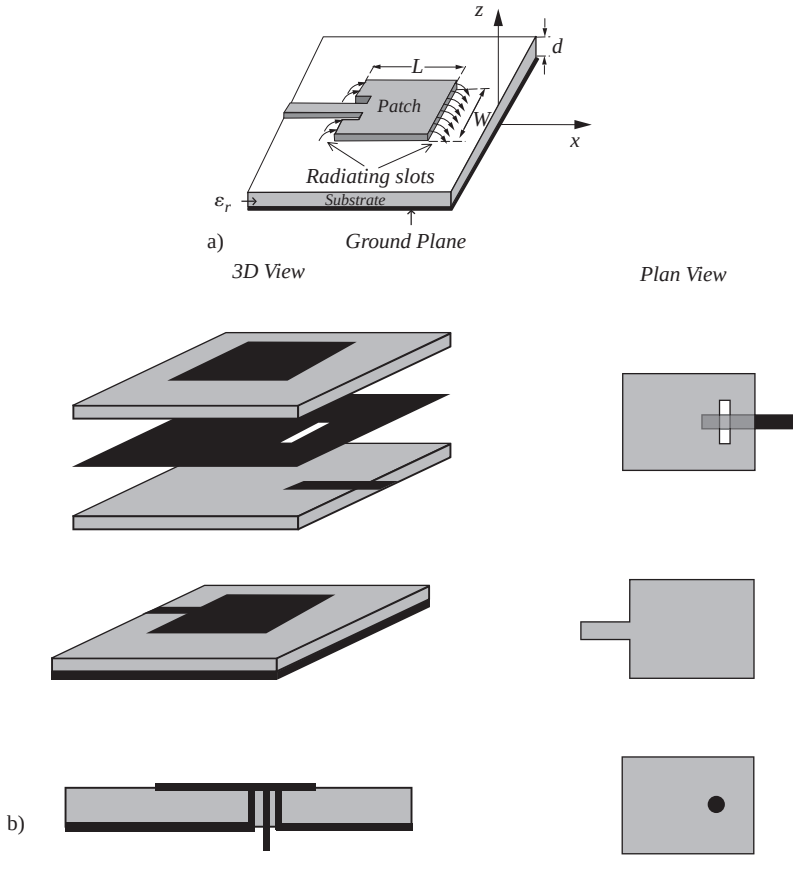


Figure 5.39 Microstrip antennas and their feeds (a) a microstrip antenna with its coordinates; (b) three feeding configurations: coupling feed, microstrip feed and coaxial feed

low power-handling capability (not suitable for high-power applications), poor polarization purity, and relatively narrow frequency bandwidth. However, significant improvements in, for example, broadening the bandwidth have been made (by stacking and other methods in order to meet the demand from the booming wireless communications industry [25]). Considerable attention has been paid recently to how to make this kind of antenna tunable and reconfigurable by adding loads/switches on or between the patch and the ground plane. The future for this antenna seems bright and exciting.

In this section a rectangular patch antenna is chosen as an example for investigation, since it is the most popular printed antenna. We are going to examine the operational principles, major characteristics and design procedures. As usual, a design example will be given at the end of the section.

5.2.5.1 Operational Principles

The rectangular antenna dimensions and coordinates are displayed in Figure 5.39(a). Usually, the patch length L is between $\lambda_0/3$ and $\lambda_0/2$ and its width W is smaller than λ_0 (it cannot

be too small, otherwise the antenna becomes a microstrip line, which is not a radiator) while its thickness t is extremely small. The substrate of thickness d ($\ll \lambda_0$) uses the same kind of materials as listed in Table 2.4. Their relative permittivity is normally between 2 and 24.

To be a resonant antenna, the length L should be around half of the wavelength. In this case, the antenna can be considered a *λ/2 transmission line resonant cavity with two open ends where the fringing fields from the patch to the ground are exposed to the upper half space ($z > 0$) and are responsible for the radiation*. This radiation mechanism is the same as the slot line, thus there are two radiating slots on a patch antenna, as indicated in Figure 5.39(a). This is why the microstrip antenna can be considered an aperture-type antenna.

The fringing fields at the ends are separated by λ/2, which means that they are 180 degrees out of phase but are equal in magnitude. Viewed from the top of the antenna, both fields are actually in phase for the x components, which leads to a broadside radiation with a maximum in the z direction.

5.2.5.2 Analysis and Design

As a resonant cavity, there are many possible modes (like waveguides), thus a patch antenna is multimode and may have many resonant frequencies. The fundamental and dominant mode is TM_{100} (a half wave change along the x-axis and no changes along the other two axes).

Radiation Pattern and Directivity

The radiation comes from the fringing fields at the two open ends, as discussed above, which is equivalent to two slot antennas separated by a distance L. It can be proved that the far-field electric field can be expressed as:

$$\boldsymbol{E} = E_0 \sin c\left(\frac{\beta W}{2}\sin\theta\sin\phi\right)\cos\left(\frac{\beta L}{2}\sin\theta\cos\phi\right)(\hat{\theta}\cos\phi - \hat{\phi}\cos\theta\sin\phi) \tag{5.85}$$

where β is the free space wave number. The first factor is the pattern factor for a uniform line source of width W in the y direction (similar to Example 5.4) and the second factor is the array factor for the two-element slots separated by L in the x direction (to be discussed in the next section). For both components, the peak is at $\theta = 0$, which corresponds to the z direction. It has a broadside unidirectional pattern.

The radiation patterns in the two principal planes are

a) E-plane ($\phi = 0^\circ$):

$$\boldsymbol{E} = \hat{\boldsymbol{\theta}} E_0 \cos\left(\frac{\beta L}{2}\sin\theta\right) \tag{5.86}$$

b) H-plane ($\phi = 90^\circ$):

$$\boldsymbol{E} = -\hat{\phi} E_0 \sin c\left(\frac{\beta W}{2}\sin\theta\right)\cos\theta \tag{5.87}$$

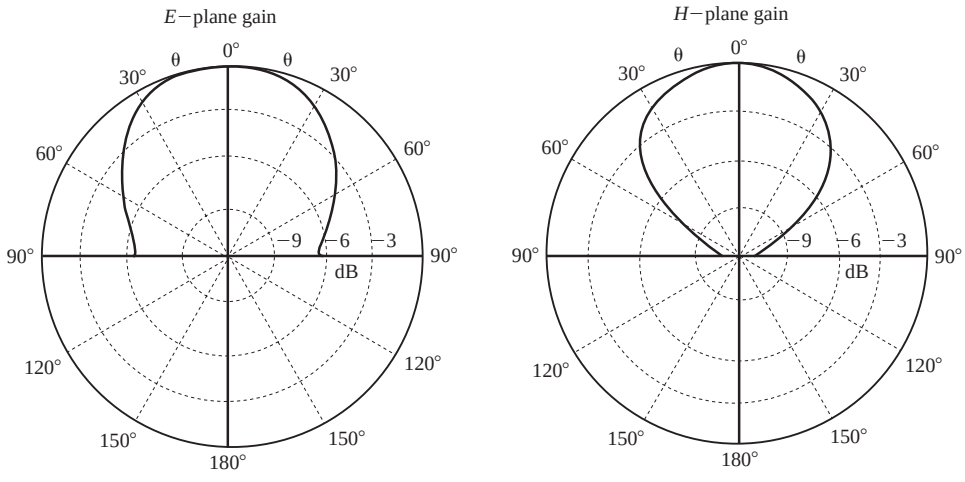


Figure 5.40 Typical radiation patterns of a resonant rectangular patch antenna

These results have neglected the substrate effects and slot width, but are good enough as estimates. The typical radiation patterns in the E- and H-planes are shown in Figure 5.40. As discussed in Section 5.1.2, if the ground plane is finite, leaky radiation towards the lower half space will occur.

With the radiated field, we are able to compute the directivity using Equation (4.11). Asymptotically, the directivity of the microstrip antenna can be expressed as

$$D = \begin{cases} 6.6 = 8.2\,\text{dBi}, & W \ll \lambda_0 \\ 8W/\lambda_{0,} & W \gg \lambda_0 \end{cases} \tag{5.88}$$

The larger the width, the larger the directivity. These are approximations as well; the actual directivities are slightly lower than these values.

Input Impedance and Bandwidth

According to Equation (40) of [23], the typical impedance at the edge of a resonant rectangular patch ranges from 100 to 400 Ω, the radiation impedance of a patch at the edge can be approximated as

$$Z_a \approx 90 \frac{\varepsilon_r^2}{\varepsilon_r - 1} \left(\frac{L}{W}\right)^2 (\Omega) \tag{5.89}$$

Here the radiation efficiency (a function of the thickness) is assumed to be 100%. Thus, the impedance is determined by three parameters. For a PTFE (Teflon) based substrate with a relative permittivity of 2.1, to obtain a 50 Ω input impedance we need $(L/W) = 0.3723$. Since $L = 0.49\lambda$, we have $W = 1.316\lambda$.

It has also been found that an empirical formula can be used to estimate the impedance fractional bandwidth for $VSWR < 2$ [23]:

$$\frac{\Delta f}{f_0} = \frac{16}{3\sqrt{2}} \frac{\varepsilon_r - 1}{\varepsilon_r^2} \frac{Ld}{\lambda W} \approx 3.77 \frac{\varepsilon_r - 1}{\varepsilon_r^2} \frac{Ld}{\lambda W} \tag{5.90}$$

Thus, the bandwidth is proportional to the thickness of the substrate. This also indicates that the higher the permittivity, the smaller the bandwidth, which means there is a trade-off between the size (Ld/W) and bandwidth.

Design Equations and Procedures

Because of the fringing effects, electrically the patch of the antenna looks larger than its physical dimensions; the enlargement on L is given by [2]

$$\Delta L = 0.412d(\varepsilon_{reff} + 0.3)(W/d + 0.264)/[(\varepsilon_{reff} - 0.258)(W/d + 0.8)] \tag{5.91}$$

Where the effective (relative) permittivity is

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2\sqrt{1 + 12d/W}} \tag{5.92}$$

This is related to the ratio of d/W. The larger the d/W, the smaller the effective permittivity.

The effective length of the patch is now

$$L_{eff} = L + 2\Delta L \tag{5.93}$$

The resonant frequency for the TM_{100} mode is

$$f_r = \frac{1}{2L_{eff}\sqrt{\varepsilon_{reff}}\sqrt{\varepsilon_0\mu_0}} = \frac{1}{2(L + 2\Delta L)\sqrt{\varepsilon_{reff}}\sqrt{\varepsilon_0\mu_0}} \tag{5.94}$$

An optimized width for an efficient radiator is

$$W = \frac{1}{2f_r\sqrt{\varepsilon_0\mu_0}}\sqrt{2/(\varepsilon_r + 1)} \tag{5.95}$$

Now the design problem: if the substrate parameters (ε_r and d) and the operational frequency (f_r) are known, how can we obtain the dimensions of the patch antenna (W and L)?

Based on these simplified formulas, we can adopt the following design procedure to design the antenna:

Step 1: *Use Equation (5.95) to find the width W.*
Step 2: *Calculate the effective permittivity ε_{reff} using Equation (5.92).*
Step 3: *Compute the extension of the length ΔL using Equation (5.91).*

Step 4: *Determine the length L by solving Equation (5.94) for L, giving the solution*

$$L = \frac{1}{2 f_r \sqrt{\varepsilon_{reff}} \sqrt{\varepsilon_0 \mu_0}} - 2\Delta L \tag{5.96}$$

Example 5.7: Design a rectangular patch. RT/Duroid 5880 substrate ($\varepsilon_r = 2.2$ and $d = 1.588$ mm) is to be used to make a resonant rectangular patch antenna of linear polarization.

a. Design such an antenna to work at 2.45 GHz for Bluetooth applications.
b. Estimate its directivity.
c. If it is to be connected to a 50 ohm microstrip using the same PCB board, design the feed to this antenna.
d. Find the fractional bandwidth for $VSWR < 2$.

Solution:

a. Follow the design procedure suggested above to obtain:

$$W = \frac{1}{2 f_r \sqrt{\varepsilon_0 \mu_0}} \sqrt{2/(\varepsilon_r + 1)} = \frac{300}{2 \times 2.45} \sqrt{2/(2.2 + 1)} = 48.40\,(\text{mm})$$

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2\sqrt{1 + 12d/W}} = \frac{2.2 + 1}{2} + \frac{2.2 - 1}{2\sqrt{1 + 12 \times 1.588/48.4}} = 2.1082$$

$$\Delta L = 0.412 \times 1.588(2.1082 + 0.3)\left(\frac{48.4}{1.588} + 0.264\right) \Big/ \left[(2.1082 - 0.258)\left(\frac{48.4}{1.588} + 0.8\right)\right] = 0.84\,(\text{mm})$$

and

$$L = \frac{1}{2 f_r \sqrt{\varepsilon_{reff}} \sqrt{\varepsilon_0 \mu_0}} - 2\Delta L = \frac{300}{2 \times 2.45\sqrt{2.1082}} - 2 \times 0.837 = 40.49\ (\text{mm})$$

Thus, the designed patch should have $L = 40.49$ mm and $W = 48.40$ mm.

b. Since the wavelength at 2.45 GHz is 122.45 mm $> W$, using Equation (5.88) gives the maximum directivity, which is about 6.6 or 8.2 dBi.

c. The input impedance is given by (0.89) and is

$$Z_a = 90 \frac{\varepsilon_r^2}{\varepsilon_r - 1} \left(\frac{L}{W}\right)^2 = 90 \frac{2.2^2}{2.2 - 1} \left(\frac{40.49}{48.40}\right)^2 = 254.04\Omega$$

which does not match well with a 50 Ω standard microstrip and therefore a quarter-wavelength transformer is used to connect them (refer to Chapter 2). The characteristic impedance of the transition section should be

$$Z_T = \sqrt{50 \times 254} = 122.69\,\Omega$$

Thus, we can use Equation (2.81) to determine the transition line width, which is:

$$Z_0 = \frac{60}{\sqrt{\varepsilon_r}} \ln\left(\frac{8d}{W_T} + \frac{W_T}{4d}\right) = \frac{60}{\sqrt{2.2}} \ln\left(\frac{8 \times 1.588}{W_T} + \frac{W_T}{4 \times 1.588}\right) = 122.69$$

Thus, $W_T = 0.615$ (mm), which is very thin. To obtain its length, we need to calculate the relative effective permittivity for the line, which is given by Equation (2.78):

$$\varepsilon_{re} \approx \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2\sqrt{1 + 12d/W_T}} \approx 1.706$$

Hence, the length of the transition should be:

$$\frac{\lambda}{4} = \frac{\lambda_0}{4\sqrt{\varepsilon_{re}}} = \frac{122.45}{4\sqrt{1.706}} = 23.437 \text{ (mm)}$$

The width of the 50 Ω microstrip feed line can be found using Equation (2.81), i.e.

$$Z_0 = \frac{120\pi}{\sqrt{\varepsilon_r}(W_m/d + 1.393 + 0.667 \ln(W_m/d + 1.44))} = 50$$

The solution is $W_m = 4.367$ mm. The final design is shown in Figure 5.41.

Since the width of the quarter-wave transformer might be too thin to make properly in practice, an alternative design is to employ an inset feed, as shown in Figure 5.39(a) and use the following design equation [2, 24, 26]:

$$R_{in}(x = x_0) = R_{in}(x = 0) \cos^2\left(\frac{\pi}{L} x_0\right) \tag{5.97}$$

We can find that the recessed distance (the length cutting into the patch) is

$$x_0 = \frac{L}{\pi} \cos^{-1} \sqrt{\frac{50}{122.69}} \approx 0.2796L = 11.32 \text{ (mm)}$$

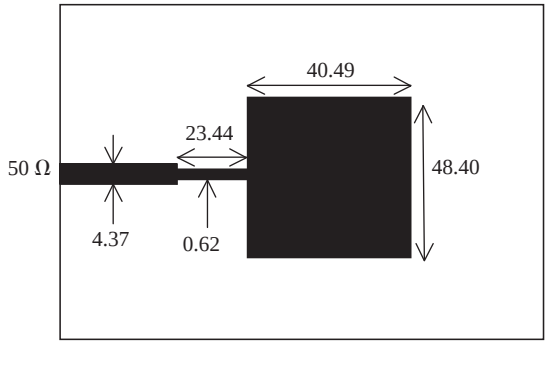


Figure 5.41 A matched resonant patch antenna for 2.45 GHz

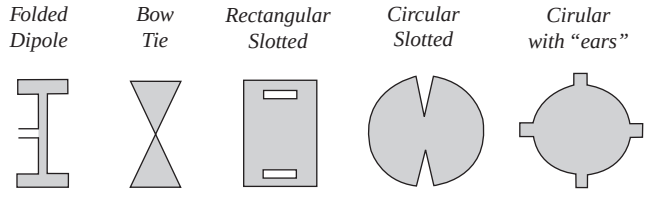


Figure 5.42 A variety of patch antennas

In this case no transformer is used and the 50 Ω line can be directly connected to the antenna with an inset of length 11.32 mm, which appears to be a good solution.

d. Using Equation (5.90), we can find the fractional bandwidth for $VSWR < 2$:

$$\frac{\Delta f}{f_0} = 3.77\frac{\varepsilon_r - 1}{\varepsilon_r^2}\frac{Ld}{\lambda W} = 3.77\frac{2.2-1}{2.2^2}\frac{1.588 \times 40.49}{122.45 \times 48.40} = 0.0101 = 1.01\%$$

The bandwidth is indeed very narrow. Some alternative patch antennas are displayed in Figure 5.42. Generally speaking, they can offer broader bandwidth than a rectangular patch.

5.2.5.3 Ground Plane

The ground plane is part of the antenna. Ideally, the ground plane should be infinite as for a monopole antenna. But, in reality, a small ground plane is desirable. As shown in Figure 5.39(a), the radiation of a microstrip antenna is generated by the fringing field between the patch and the ground plane, the minimum size of the ground plane is therefore related to the thickness of the dielectric substrate. Generally speaking, a λ/4 extension from the edge of the patch is required for the ground plane, whereas the radius of a monopole ground plane should be at least one wavelength.

5.3 Antenna Arrays

So far we have studied many different antennas: wire types and aperture types; resonant antennas (such as dipole and patch antennas) and traveling wave antennas (such as Yagi–Uda and periodic antennas). They can all be classified as single-element antennas. Once the frequency is given, everything (the radiation pattern, input impedance, etc.) is fixed. They lack flexibility and the gain is normally very limited. A high-gain antenna means that the aperture size of the antenna has to be very large, which may be a problem in practice. Also, sometimes we need to be able to control the antenna radiation pattern, for example for tracking or anti-jamming/interference applications. A single-element antenna is not good enough to meet such a requirement. In this case, an antenna array could be a good solution.

An *antenna array* consists of more than one antenna element and these radiating elements are strategically placed in space to form an array with desired characteristics, which are achieved by varying the feed (amplitude and phase) and relative position of each radiating element. The total radiated field is determined by vector addition of the fields radiated by the individual

elements. The total dimensions of the antenna are enlarged without increasing the size of the individual element. The major advantages of an array are:

- the flexibility to form a desired radiation pattern;
- the high directivity and gain;
- the ability to provide an electrically scanned beam (mechanical rotation can be avoided).

The main drawbacks are:

- the complexity of the feeding network required;
- the bandwidth limitation (mainly due to the feeding network).

In this section we are going to examine the basic operational principle of an antenna array and its associated theory, how to design an array and what the problems and trade-offs are.

5.3.1 Basic Concept

Let us assume that there are N elements in an antenna array, as shown in Figure 5.43. The phase and amplitude of each element can be tuned electrically or mechanically using phase shifters and attenuators. Sometimes, these weighting factors are already integrated into the radiating elements and no tuning is required.

The total radiated field can be obtained by summing up the radiated field from each element antenna, i.e.

$$\boldsymbol{E}(r,\theta,\phi) = \sum_{n=1}^{N} \boldsymbol{E_n}(J_n) \tag{5.98}$$

Each element has many variables, including the element antenna itself and its excitation (amplitude and phase); the overall arrangement of these elements is another variable which can be

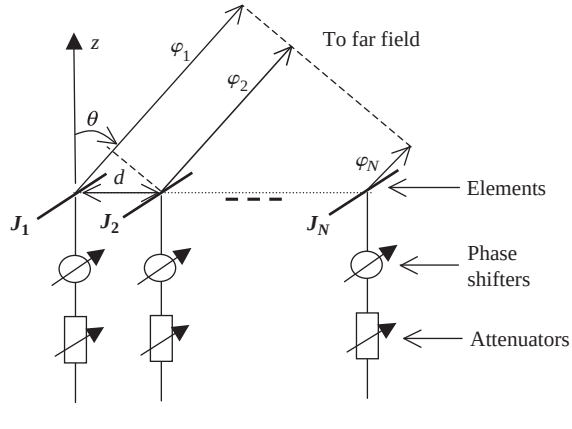


Figure 5.43 A typical antenna array of N elements

in many different geometrical forms to achieve the desired radiation characteristics. In practice, all the elements are placed in a specific configuration to form a well-controlled 1D, 2D or even 3D array (which may be linear, circular, rectangular or elliptical in shape). Arbitrary shaped arrays are not well studied and the patterns are not easy to control, thus they are not used.

5.3.2 Isotropic Linear Arrays

Now we are going to use a linear array, the most popular array arrangement, as an example to explore the features of an antenna array. To simplify the analysis, the radiator is assumed to be an isotropic element but its amplitude and phase are controllable. Since the isotropic element means that its radiation is the same in every direction, from Equation (4.6), we know that the far-field radiation from such an isotropic element can be expressed as

$$E_{ie} = \frac{e^{-j\beta r}}{4\pi r} \tag{5.99}$$

The polarization is not shown here but the field is orthogonal to the propagation direction. Thus, the total radiated field from an isotropic linear array of N elements can be written as

$$E_{ia} = \sum_{n=1}^{N} E_{ie} A_n e^{j\varphi_n} = \frac{e^{-j\beta r}}{4\pi r} \sum_{n=1}^{N} A_n e^{j\varphi_n} = E_{ie} \sum_{n=1}^{N} A_n e^{j\varphi_n} \tag{5.100}$$

where A_n is the amplitude and φ_n is the relative phase (to a common reference, as shown in Figure 5.43) of element n. The second term of Equation (5.100) is called the (isotropic) *array factor*, that is

$$AF = \sum_{n=1}^{N} A_n e^{j\varphi_n} \tag{5.101}$$

If we further assume that the spacing between elements is a constant d and all the elements have identical amplitude (say, 1) but each succeeding element has a φ_0 progressive phase lead current excitation relative to the preceding one, this array is referred to as a *uniform array*. Using Figure 5.43, the array factor in this case is:

$$\begin{aligned} AF &= \sum_{n=1}^{N} A_n e^{j\varphi_n} = 1 + e^{j(\beta d \sin\theta + \varphi_0)} + e^{j2(\beta d \sin\theta + \varphi_0)} + \cdots + e^{j(N-1)(\beta d \sin\theta + \varphi_0)} \\ &= \sum_{n=1}^{N} e^{j(n-1)(\beta d \sin\theta + \varphi_0)} = \sum_{n=1}^{N} e^{j(n-1)\Psi} \end{aligned} \tag{5.102}$$

where $\Psi = \beta d \sin\theta + \varphi_0$ is the phase difference (lead) between two adjacent elements and the first element is chosen as the reference antenna. The geometric series can be simplified to:

$$AF = \frac{1 - e^{jN\Psi}}{1 - e^{j\Psi}} = e^{j[(N-1)/2]\Psi} \left[\frac{\sin(N\Psi/2)}{\sin(\Psi/2)} \right] \tag{5.103}$$

The maximum of the above AF is N, the normalized antenna factor of a uniform array is therefore

$$AF_n = \frac{1}{N}\left[\frac{\sin(N\Psi/2)}{\sin(\Psi/2)}\right] \tag{5.104}$$

The normalized AF_n in dB as a function of θ in degrees for $\varphi_0 = 0$ is plotted in Figure 5.44 for (a) $N = 20$ and $d = \lambda$, (b) $N = 10$ and $d = \lambda$ and (c) $N = 10$ and $d = \lambda/2$. The maximum at $\theta = 0^\circ$ is the main beam or main lobe. The other maxima are called grating lobes. Generally, grating lobes are undesirable, because most applications such as radars require a single, focused beam. The important messages from these plots and Equation 5.104 are:

- there are $(N - 2)$ side lobes between the main and first grating lobes (not applicable for small d, to be discussed in Figure 5.47);
- grating lobes appear if $d \geq \lambda$. To avoid any grating lobes, the largest separation between the element should be less than one wavelength;
- the notches between lobes are nulls (in theory);
- the first side lobe is about -13 dB, the same as for uniform aperture antennas;

5.3.2.1 Phased Arrays

From Equation (5.104), we know that the maximum of the radiation occurs at $\psi = 0$ which is

$$\beta d \sin\theta + \varphi_0 = 0 \tag{5.105}$$

or at the angle

$$\theta = -\sin^{-1}\left(\frac{\lambda\varphi_0}{2\pi d}\right) \tag{5.106}$$

Normally, the spacing d is fixed for an array; we can control the maximum radiation (or scan the beam) by changing the phase φ_0 and the wavelength (frequency) – this is the principle of a *phase/frequency scanned array*. The phase change is accomplished electronically by the use of ferrite (sensitive to magnetic field) or diode (sensitive to bias voltage) phase shifters, switched transmission lines are also commonly used. Controlling the beam is one of the most appealing features of arrays. The pattern can be scanned electronically by adjusting the array coefficients (φ_0 or λ). An example of the array pattern (AF) for $N = 10$, $d = \lambda/2$ and $\varphi_0 = 45^\circ$ is shown in Figure 5.45. It is apparent that the maximum is now shifted from 0° to another angle ($\theta = -45^\circ/\pi \approx 14.3^\circ$). The angle is controlled by the phase shifter. This approach permits beams to be moved from point to point in space in just microseconds, whereas mechanical scanning of a large antenna could take several seconds or longer.

Phased antenna arrays (1D and 2D) are widely used, especially for military applications. The benefits are obvious, so are the difficulties at the feeding network. Just imagine the wiring for an antenna array with hundreds of elements!

Due to the feasibility and variables within a phased array, it is also known as a *smart antenna*.

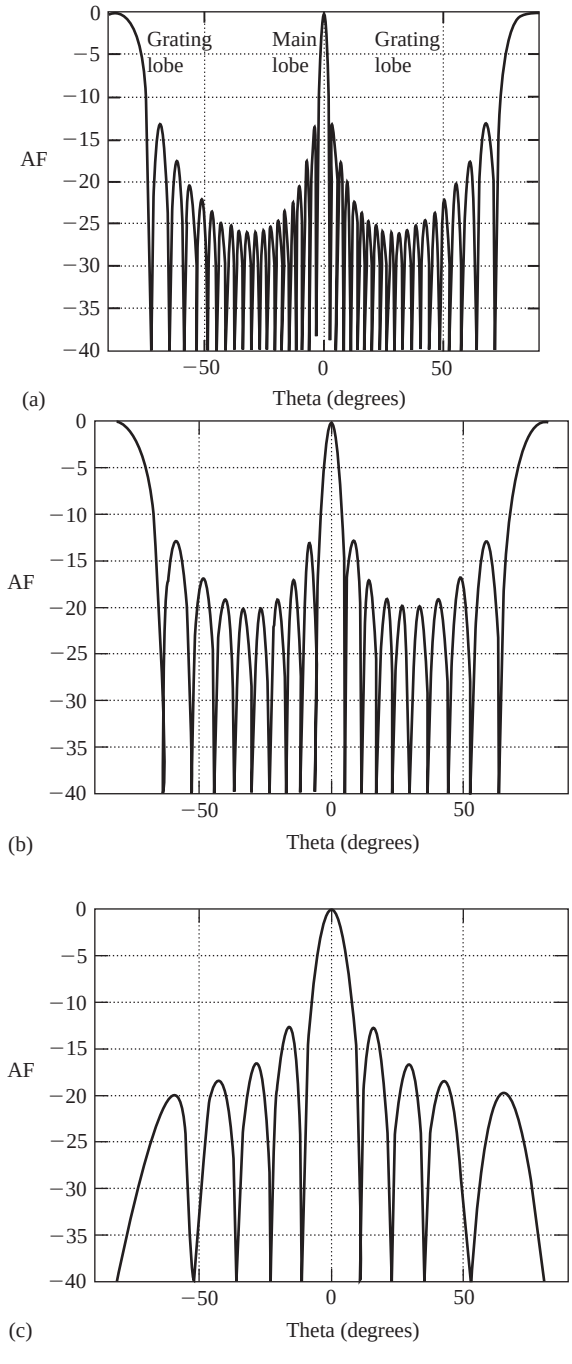


Figure 5.44 Antenna factors as a function of the scan angle θ for $\varphi_0 = 0$ (a) $N = 20$ and $d = \lambda$; (b) $N = 10$ and $d = \lambda$; (c) $N = 10$ and $d = \lambda/2$

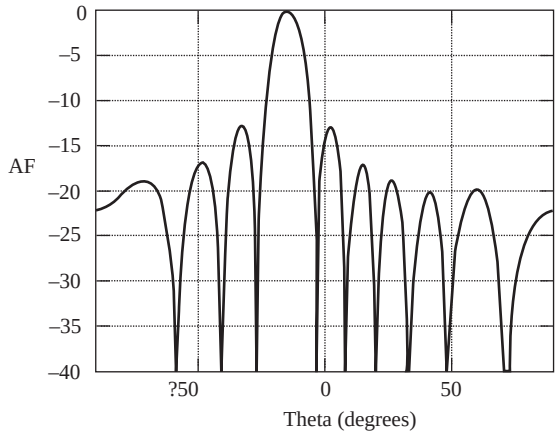


Figure 5.45 Antenna factor as a function of angle θ for $N = 10$ and $d = \lambda/2$, and $\varphi_0 = 45^\circ$

5.3.2.2 Broadside and End-fire Arrays

An array is called a *broadside array* if the maximum radiation of the array is directed normal to its axis ($\theta = 0^\circ$ in Figure 5.43); while it is called an *end-fire array* if the maximum radiation is directed along the axis of the array ($\theta = 90^\circ$ in Figure 5.43). For a given frequency, we can control the spacing d and the progressive phase φ_0 to make the array broadside or end-fire. The specific conditions are:

- broadside: $\varphi_0 = 0$;
- end-fire: $\varphi_0 = \pm 2\pi d/\lambda$.

The major characteristics of the radiation pattern of both arrays are listed in Table 5.6. The polar and rectangular plots of the radiation patterns of the broadside and end-fire arrays for $d = \lambda/2$ and $\lambda/4$ at N = 10 are shown in Figures 5.46 and 5.47, respectively.

It should be noted that from these results, for the same geometrical configuration (length and spacing),

- the HPBW of the end-fire array is always larger than that of the broadside array, so is the FNBW;
- the directivity of the end-fire array, as given by Table 5.6, is twice that for the broadside array if the spacing is much smaller than the wavelength.

These two conclusions seem to contradict the relation between the directivity and HPBW shown in Equation (4.13). The reality is that broadside arrays can only produce a fan beam (two symmetrical main lobes) while end-fire arrays can generate a single pencil beam (just one main lobe) when

$$d \leq \frac{\lambda}{2}\left(1 - \frac{1}{2N}\right) \tag{5.107}$$

Table 5.6 Comparison of the broadside and end-fire antenna arrays

Item	Broadside	End-fire
φ_0	0	$\pm 2\pi d/\lambda$
Maxima ($\theta =$)	$\sin^{-1}\left(\pm\dfrac{m\lambda}{d}\right)$, $m = 0, 1, 2, \ldots$	$\sin^{-1}\left(1-\dfrac{m\lambda}{d}\right)$ $m = 0, 1, 2, \ldots$
HPBW	$\approx 2\left[\dfrac{\pi}{2}-\cos^{-1}\left(\dfrac{1.391\lambda}{\pi Nd}\right)\right]$	$\approx 2\cos^{-1}\left(1-\dfrac{1.391\lambda}{\pi Nd}\right)$
1st null beamwidth (FNBW)	$2\left[\dfrac{\pi}{2}-\cos^{-1}\left(\dfrac{\lambda}{Nd}\right)\right]$	$2\cos^{-1}\left(1-\dfrac{\lambda}{Nd}\right)$
Directivity	$\approx 2\dfrac{Nd}{\lambda}, \quad d < \lambda$	$\approx 4\dfrac{Nd}{\lambda}, \quad d < \dfrac{\lambda}{2}\left(1-\dfrac{1}{2N}\right)$

Figure 5.47 (b) is an example to demonstrate this interesting feature, thus Equation (4.13) cannot be applied.

It should also be pointed out that the number of side lobes for the broadside case is no longer $(N-2)$ when the spacing d becomes small, as shown in Figure 5.47(a). Some side lobes are merged.

5.3.2.3 The Hansen–Woodyard End-fire Array

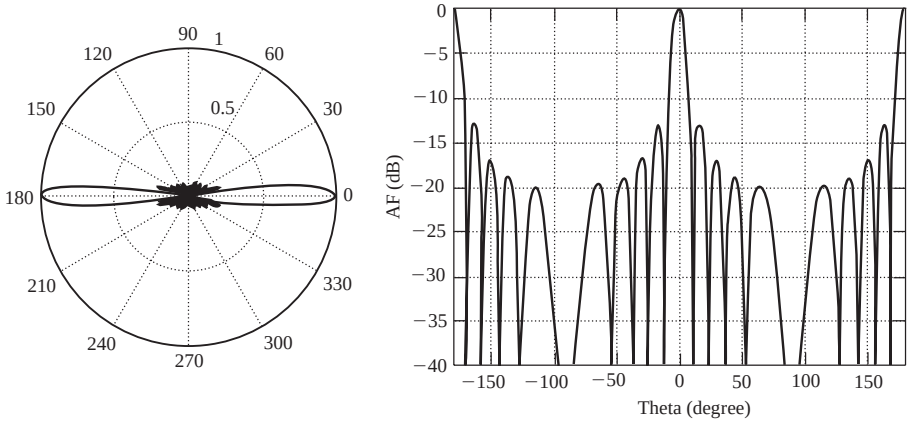
The end-fire arrays discussed above, which are called ordinary end-fire arrays, are not optimized to produce the optimum directivity. It is possible to make the main beam narrower and thus increase directivity by changing the inter-element phase shift. Hansen and Woodyard [28] found that the maximum directivity of an end-fire array is obtained when

$$\varphi_0 = \pm(2\pi d/\lambda + 2.92/N) \approx \pm(2\pi d/\lambda + \pi/N) \tag{5.108}$$

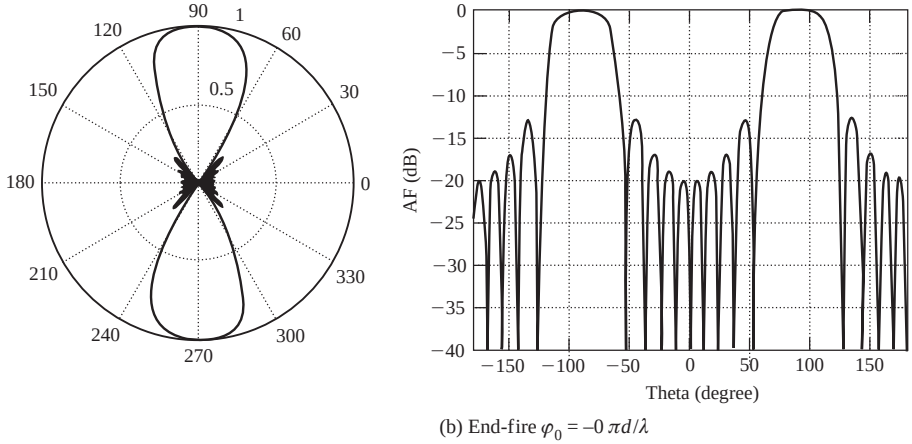
if the array length is much larger than a wavelength. However, it should be pointed out that the condition in Equation (5.108) is not for the maximum possible directivity of any array (which may occur at other directions rather than $\theta = \pm 90^\circ$) but only for end-fire arrays. The directivity can now be approximated as

$$D \approx 7.28\frac{Nd}{\lambda} \tag{5.109}$$

This means an increase of 2.6 dB over the ordinary end-fire array in directivity. This is why the *Hansen–Woodyard array* is also known as the *increased directivity end-fire array*. When the axial mode helix was discussed in Section 5.1.4, it was concluded that this type of antenna approximately satisfies the Hansen–Woodyard end-fire condition and increased directivity was obtained.



(a) Broadside $\varphi_0 = 0$

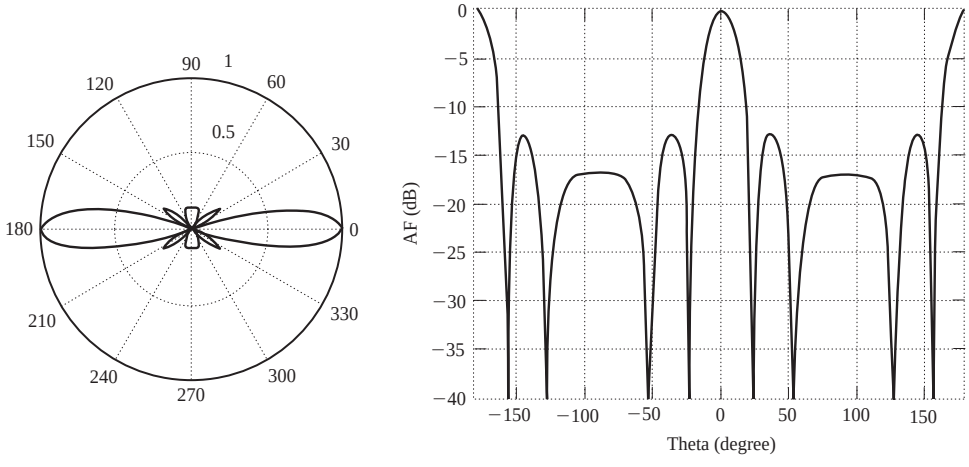


(b) End-fire $\varphi_0 = -0\ \pi d/\lambda$

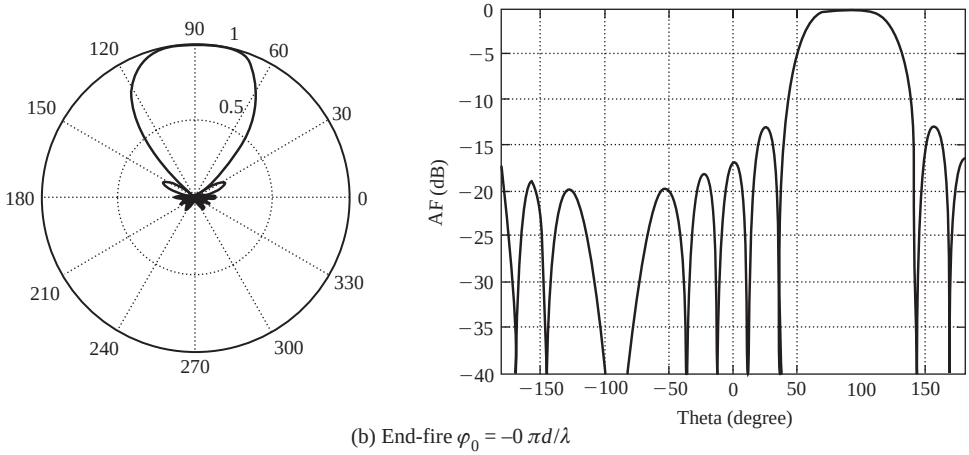
Figure 5.46 Antenna factors for broadside and end-fire arrays for $N = 10$ and $d = \lambda/2$

5.3.2.4 The Dolph–Tchebyscheff (D–T) Optimum Distribution

In addition to the directivity/gain, another very important consideration for antenna arrays is the side-lobe level (SLL), since it affects antenna performance (such as the antenna temperature and SNR). As mentioned earlier and shown in Figure 5.48, there is a trade-off between the SLL and gain. To achieve a high gain, a uniform distribution is preferred but it results in a high SLL (about -13 dB) while a well-tapered distribution can produce very low SLL or even no side lobes (such as the binominal distribution) but it exhibits a lower gain. It is shown that the far-field pattern of a linear in-phase ($\varphi_0 = 0$, broadside) array of isotropic point sources can be expressed as a finite Fourier series of N terms. We can use Dolph's procedure [29] to match the Fourier polynomial with the terms of like degree of a Tchebyscheff polynomial. This then yields the optimum source amplitude distribution for a specific SLL with all side lobes of the same level. More details can be found in [4].



(a) Broadside $\varphi_0 = 0$



(b) End-fire $\varphi_0 = -0\ \pi d/\lambda$

Figure 5.47 Antenna factors for broadside and end-fire arrays for $N = 10$ and $d = \lambda/4$

5.3.3 Pattern Multiplication Principle

In practice, isotropic sources will be replaced by practical antennas, which means that the element radiated E_{ie} in Equation (5.100) should be replaced by the radiated field of the antenna element E_e, thus the total radiated field of the antenna array is

$$E_a = E_e \sum_{n=1}^{N} A_n e^{j\varphi_n} = E_e \cdot AF \tag{5.110}$$

This is actually called the *principle of pattern multiplication*. The radiation pattern of an array is the product of the pattern of the individual element antenna with the (isotropic source) array pattern.

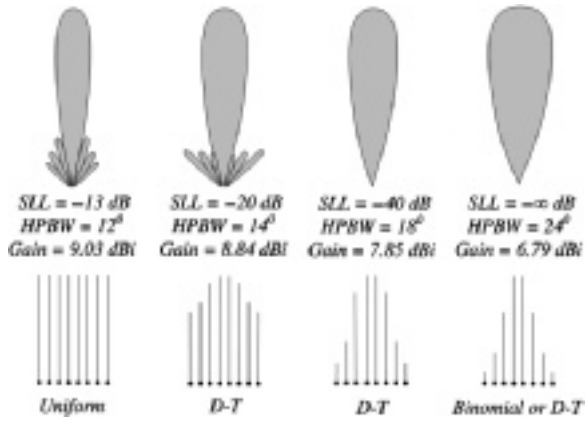


Figure 5.48 The radiation pattern (in half space), SLL, HPBW and gain for four different source distributions of eight in-phase isotropic sources spaced by $\lambda/2$. [Reproduced by Permission of The McGraw-Hill Companies]

Let us take a two-element short dipole array as an example. The elements are in phase and separated by half a wavelength, the *AF* (isotropic array pattern) is therefore a figure 8 (in shape) from Equation (5.104). Since the dipole radiation pattern is well known to us now and is also a figure 8 in shape, the total radiation patterns for the horizontally oriented and vertically oriented dipole arrays can be obtained using the pattern multiplication principle, as shown in Figure 5.49.

Using the pattern multiplication principle, we can extend the knowledge and results obtained for isotropic arrays to other antenna arrays. It is now also possible to produce a unidirectional broadside radiation pattern by using directional antenna elements.

The product of the directivity of a single element and the directivity of the isotropic array may be used to estimate the directivity of the antenna array, but the accuracy varies with the array geometry [30].

5.3.4 Element Mutual Coupling

When antenna elements are placed in an array, they interact with each other. This interaction between elements due to their close proximity is called *mutual coupling*, which affects the current distribution and hence the input impedance as well as the radiation pattern.

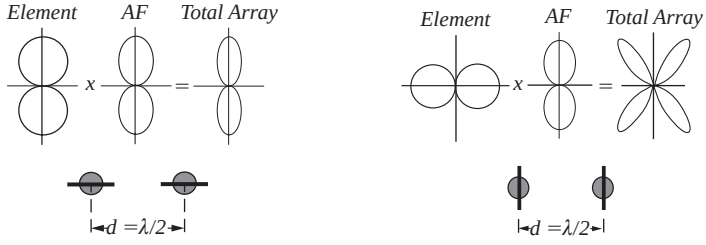


Figure 5.49 Array radiation patterns of two short dipoles separated by $d = \lambda/2$

The most significant effects are on the antenna input impedance, which is a very important parameter for a single antenna as well as for an antenna array. We can use circuit concepts, which are simpler than field concepts, to analyze the input impedance of element antennas in an array. Each element antenna is considered a current source. An antenna array of N elements is then treated as an N-port network, and the voltage generated at each element can be expressed as

$$\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2 + \cdots + Z_{1N}I_N \\ V_2 &= Z_{21}I_1 + Z_{22}I_2 + \cdots + Z_{2N}I_N \\ &\cdots \\ V_N &= Z_{N1}I_1 + Z_{N2}I_2 + \cdots + Z_{NN}I_N \end{aligned} \tag{5.111}$$

where I_n is the current at the nth element and

$$Z_{nn} = \frac{V_n}{I_n}, n = 1, 2, \ldots N \tag{5.112}$$

is called the *self-impedance* of the nth element when all other elements are open-circuited, or the input impedance when this element is isolated from the other elements (as if it were not in an array). For a half-wavelength dipole, its self-impedance is about 73 ohms. The impedance

$$Z_{mn} = \frac{V_m}{I_n} \tag{5.113}$$

is called the *mutual impedance* between the elements m and n($Z_{mn} = Z_{nm}$ by the reciprocity principle to be discussed later). Again, using two half-wavelength dipole antennas as an example, the mutual impedance (resistance and reactance) of two parallel side-by-side dipoles is shown in Figure 5.50, while the mutual impedance of two collinear dipoles as a function of

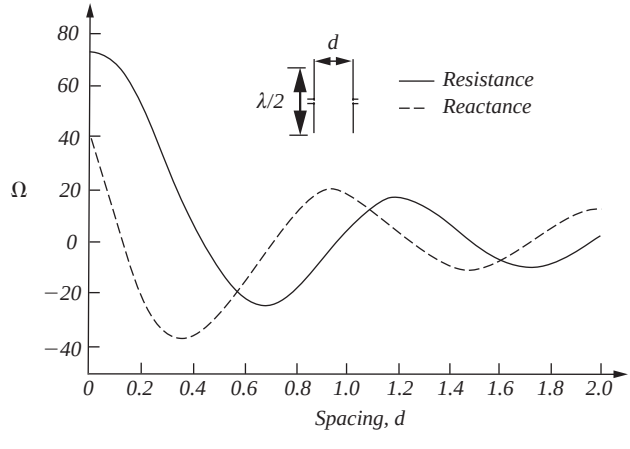


Figure 5.50 Mutual resistance and reactance of two parallel dipoles as a function of the spacing d in wavelengths. (J. D. Kraus and R. J. Marhefka, Antennas for All Applications, 3rd edition, McGraw-Hill, Inc., 2002. Reproduced by permission of The McGraw-Hill Companies)

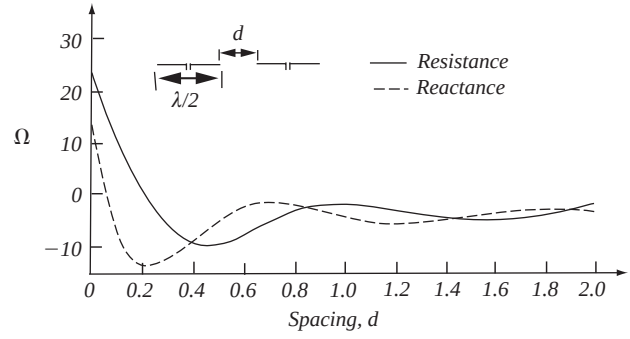


Figure 5.51 Mutual resistance and reactance of two collinear dipoles as a function of the spacing d in wavelengths. (J. D. Kraus and R. J. Marhefka, Antennas for All Applications, 3rd edition, McGraw-Hill, Inc., 2002. Reproduced by permission of The McGraw-Hill Companies)

the separation between them is plotted in Figure 5.51. Examining these results reveals that:

- the strength of the coupling decreases as spacing increases;
- the parallel configuration couples more than the collinear configuration;
- the far-field pattern of each element predicts coupling strength. The more radiation, the stronger the coupling.

Taking the mutual coupling into account, the input impedance of the nth element in the array is

$$Z_n = \frac{V_n}{I_n} = Z_{11}\frac{I_1}{I_n} + Z_{12}\frac{I_2}{I_n} + \cdots + Z_{1N}\frac{I_N}{I_n} \tag{5.114}$$

This is the actual input impedance of the element and is also referred to as the *active impedance*.

It should be pointed out that the elements at the array edges have fewer neighbouring elements, thus the mutual coupling could vary significantly from edge elements to other elements.

5.3.4.1 Discussions

In addition to 1D arrays, 2D antenna arrays have also been developed for practical applications. The slot waveguide array in Figure 5.38 is an example where each element (slot) is carefully designed to ensure the aperture field distribution on the antenna follows the desired pattern (such as a Gaussian distribution) and the impedance (including mutual impedance) is matched with the waveguide. Here discrete antenna elements are employed to form a continuous aperture field distribution. It is evident that antenna arrays are closely linked to aperture antennas. Some theories developed for aperture antennas may be applied to antenna arrays.

The array feeding network can be very complicated, thus it is a special topic. Generally speaking, just like elements in an electric circuit, the antenna elements in an array may be fed by a parallel circuit, a series circuit or a combination of both. But care has to be taken; the feeding network is normally a distributed system at high frequencies. The network may

limit the bandwidth of the antenna system and may actually act as a radiator if not properly designed!

5.4 Some Practical Considerations

Up to now, we have covered the most popular antennas: wire-type and aperture-type antennas, element and array antennas. You should have gained a good understanding of their operational principles, major field and circuit characteristics and design procedures. However, there are many practical considerations we have not yet studied, and it is not possible to deal with all aspects of antenna design and characterization in this book. In this section we are going to look at some important considerations in practice, which include:

- the differences between transmitting and receiving antennas;
- antenna feeding and matching;
- polarization;
- radomes, housings and supporting structures.

5.4.1 Transmitting and Receiving Antennas: Reciprocity

So far we have not distinguished between antennas for transmitting and receiving purposes. Are there any differences between them? The short answer is yes and no, depending on what parameters you are talking about. Let us examine these two cases from both the field point of view and the circuit point of view.

Let us assume within a linear and isotropic medium that there exist two sets of sources ($\boldsymbol{J}_1$, $\boldsymbol{M}_1$) and ($\boldsymbol{J}_2$, $\boldsymbol{M}_2$) which are allowed to radiate in the medium and produce fields ($\boldsymbol{E}_1$, $\boldsymbol{H}_1$) and ($\boldsymbol{E}_2$, $\boldsymbol{H}_2$), respectively. Using Maxwell's equations, it can be shown that [31]:

$$\int_V (\boldsymbol{E_1} \bullet \boldsymbol{J_2} - \boldsymbol{H_1} \bullet \boldsymbol{M_2}) dv = \int_V (\boldsymbol{E}_2 \bullet \boldsymbol{J}_1 - \boldsymbol{H}_2 \bullet \boldsymbol{M}_1) dv \tag{5.115}$$

This is called the *Lorentz reciprocity theorem*. This can be interpreted to mean that the coupling between a set of fields and a set of sources which produces another set of fields is the same as the coupling between another set of fields and another set of sources. We can represent this simply as

$$\langle 1, 2 \rangle = \langle 2, 1 \rangle \tag{5.116}$$

Under this condition, from the circuit point of view, the *mutual coupling impedances between two antennas are the same*

$$Z_{12} = Z_{21} \tag{5.117}$$

The self-impedance, i.e. the antenna input impedance, is also the same for the transmitting or receiving mode.

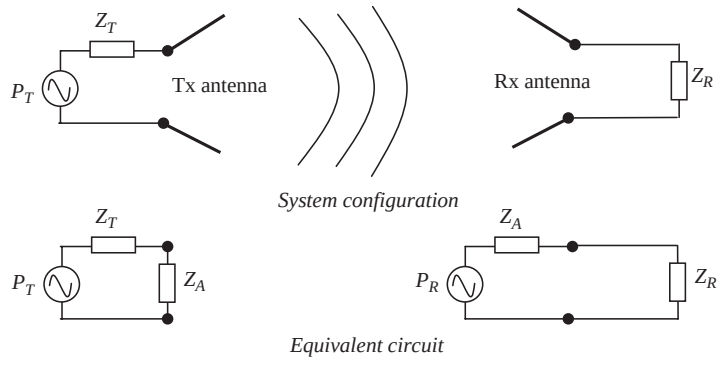


Figure 5.52 Transmitting and receiving systems and their equivalent circuits

From the field point of view, the same conclusion can be arrived at; that is, *the radiated field patterns are the same when the antenna is employed for transmitting or receiving*.

Figure 5.52 shows the system configuration for a typical transmitting and receiving system. At the transmitting side, the transmitting antenna is just a load (Z_A) to the feed line, as discussed in Section 2.3.2 (impedance matching), the antenna impedance and source internal impedance should meet the following condition to achieve the maximum transmission:

$$Z_A{}^* = Z_T \text{ or } R_A - jX_A = R_T + jX_T.$$

The power radiated is given by

$$P_{rad} = \frac{R_A}{R_T + R_A} P_T \tag{5.118}$$

While at the receiving side, the receiving antenna is equivalent to a source with internal impedance Z_A,as shown in Figure 5.52, which is different from the transmitting antenna. Similarly, to reach the maximum power transfer, the antenna impedance and the receiver impedance should meet the following condition:

$$Z_A{}^* = Z_R \text{ or } R_A - jX_A = R_R + jX_R$$

and the power received by the receiver is given by

$$P_{rec} = \frac{R_R}{R_A + R_R} P_R \tag{5.119}$$

For a perfect match, $R_A = R_R$, we have

$$P_{rec} = 0.5P_R \leq 0.5\frac{R_A}{R_T + R_A} P_T \tag{5.120}$$

This means that the receiver gets half of the power collected by the receiving antenna and the other half is reradiated. *If both the transmitting and receiving sides are perfectly matched and*

there is no propagation pathloss, the receiver can only yield 25% of the source power P_T. This percentage share can be increased to a maximum of 50% by increasing the radiating efficiency $\frac{R_A}{R_T+R_A}$ towards 100%. For example, if $R_A = 300\,\Omega$ and $R_T = 50\,\Omega$, $\frac{R_A}{R_T+R_A} \approx 85.7\% > 50\%$, and the maximum possible received power is $P_{rec} = 0.5P_R = 0.5\frac{R_A}{R_T+R_A}P_T \approx 0.43P_T$, which is much greater than 25% of the source power. It should be pointed out that the mismatch efficiency of the antenna to feed line, which was discussed in Section 4.2.1, has not been taken into account here; this could become an issue at high frequencies.

From a system design point of view, the requirements on the transmitting antenna and the receiving antenna are not the same. For transmitting, the paramount importance is to transmit enough power to the desired directions, thus the maximum radiating efficiency and antenna gain are the most important parameters. For receiving, the most important requirement is usually a large signal-to-noise ratio (SNR). Although the efficiency and gain are still important, low side lobes may be more desirable than other parameters. Another thing is that the transient responses of transmitting and receiving antennas are different: an (small) antenna acts as a differentiator on transmit but a voltage probe (replica of the incident field) on receive.

It is now clear that antenna properties (such as the radiation pattern and input impedance) are basically the same whether it is used as a transmitting antenna or a receiving antenna. Thus, there is no need to distinguish it as a transmitting or receiving antenna. However, the design requirements for the transmitting antenna and receiving antenna are usually different in practice. Thus, the design considerations for a transmitting antenna and for a receiving antenna may be different.

5.4.2 Baluns and Impedance Matching

As discussed earlier, an antenna is normally connected to a transmission line and good matching between them is very important. A coaxial cable is often employed to connect an antenna–mainly due to its good performance and low cost. A half-wavelength dipole antenna with impedance of about 73 ohms is widely used in practice. From the impedance-matching point of view, this dipole can match well with a 50 or 75 ohm standard coaxial cable. Now the question is: can we connect a coaxial cable directly to a dipole?

As illustrated in Figure 5.53, when a dipole is directly connected to a coaxial cable, there is a problem: a part of the current coming from the outer conductor of the cable may go to the outside of the outer conductor at the end and return to the source rather than flow to the dipole. This undesirable current will make the cable become part of the antenna and radiate or receive

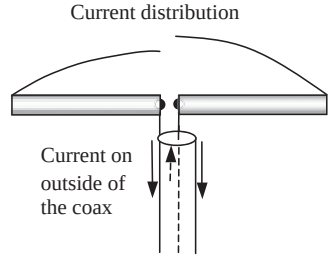


Figure 5.53 A dipole connected to a coax

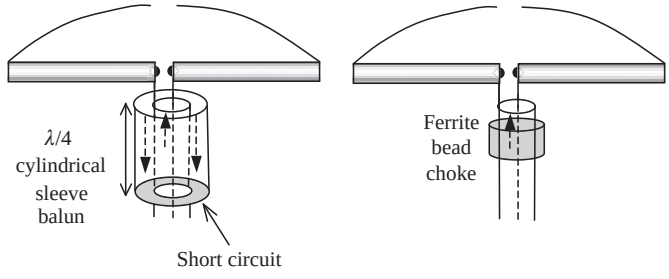


Figure 5.54 Two examples of baluns

unwanted signals, which could be a very serious problem in some cases. In order to resolve this problem, a balun is required.

The term *balun* is an abbreviation of the two words *bal*ance and *un*balance. It is a device that connects a balanced antenna (a dipole, in this case) to an unbalanced transmission line (a coaxial cable, where the inner conductor is not balanced with the outer conductor). The aim is to eliminate the undesirable current coming back on the outside of the cable. There are a few baluns developed for this important application. Figure 5.54 shows two examples. The sleeve balun is a very compact configuration: a metal tube of $\lambda/4$ is added to the cable to form another transmission line (a coax again) with the outer conductor cable, and a short circuit is made at the base which produces an infinite impedance at the open top. The leaky current is reflected back with a phase shift of 180 degrees, which results in the cancellation of the unwanted current on the outside of the cable. This balun is a narrowband device. If the short-circuit end is made as a sliding bar, it can be adjusted for a wide frequency range (but it is still a narrowband device). The second example in Figure 5.54 is a ferrite-bead choke placed on the outside of the coaxial cable. It is widely used in the EMC industry and its function is to produce a high impedance, more precisely high inductance due to the large permeability of the ferrite, and act as a filter. With good-quality ferrite beads, a large bandwidth may be obtained (an octave or more). But this device is normally just suitable for the frequencies below 1 GHz, which is determined by the ferrite properties. It should be pointed out that a choke is not usually termed a balun – also, at high frequencies it can be lossy, which can reduce the measured antenna efficiency.

There are some other types of baluns; a good discussion can be found in [4]. For example, the tapered balun (see the feed to the TEM horn antenna in Figure 5.23) is broadband but its length is usually very long ($> \lambda/2$).

When combined with impedance-matching techniques, a balun can become an impedance transformer (as well as a balun) [2]. The most well-known examples are the $\lambda/2$ coaxial balun transformer (4:1) and ferrite core transformers (n:1). To connect a folded dipole (shown in Figure 5.4) of impedance about 280 ohms to a coaxial cable of 50 ohms, a $\lambda/2$ coaxial balun transformer could be used to obtain a good impedance match and balanced feed. A general discussion on impedance-matching techniques was given in Section 2.3.2.

5.4.3 Antenna Polarization

Radio wave polarization was introduced in Section 3.2.2. There are linearly, circularly and elliptically polarized waves. Antennas are the devices used to generate and receive these waves. But how is the antenna polarization linked to the wave polarization?

Figure 5.55 Four antennas for circular polarization (a) two orthogonal dipoles; (b) a dipole and a loop; (c) a slotted cylinder; (d) a circular patch

The antenna polarization is determined by the polarization of its radiating wave. For example, a dipole is a linearly polarized antenna since its radiating wave has just one component (E_θ). This is relatively easy to understand. The difficulty comes with circularly polarized waves. How does one generate a right-hand or left-hand circularly polarized wave?

5.4.3.1 Circular Polarization

We recall, from Equations (3.22) and (3.23), that a circularly polarized wave must have two orthogonal electric field components. They have the same amplitude but a phase difference of 90 degrees. So far we have only discussed one circularly polarized antenna: the helical antenna, which is also the most popular circularly polarized antenna. It produces two orthogonal components E_θ and E_ϕ (given by Equations (5.21) and (5.22)), which are 90 degrees out of phase.

There are many other antenna designs which can produce circular polarization. Four simple examples are shown in Figure 5.55, where (a) shows two orthogonal dipoles which are fed with a 90-degree phase difference (this can be generated by a phase shifter or $\lambda/4$ physical separation). Which dipole is leading in phase determines the rotation of the circular wave, LCP or RCP; if the radiation in one axial direction is LCP, it is RCP in the opposite axial direction. Figure 5.55(b) shows a combination of a dipole and a loop, which are also of 90-degree phase difference; (c) shows a slotted dipole cylinder and their feeds are also 90 degrees out of phase; and (d) is a circular patched antenna with two symmetrical cuts which ensure that two orthogonal currents are generated with a 90-degree phase difference. The first three configurations can provide omnidirectional patterns.

Some of the arrangements (such as the helical antenna or design (a) in Figure 5.55) can be placed within a circular waveguide so as to generate a circular TE_{11} mode and a circular polarized beam can be produced when the waveguide is flared into a conical horn antenna. Some of them can serve as a primary antenna of a reflector antenna so as to produce highly directional and circular radiation.

5.4.3.2 Polarization Match and Mismatch

For a good radio communication system, the polarization must be matched. That is, a linearly polarized wave can be received by a linearly co-polarized (not cross-polarized) antenna, and similarly, a circularly polarized wave should be received by a circularly co-polarized antenna. They must be of the same polarization, otherwise some or even all the received power could be lost due to polarization mismatch. For example, if an incoming wave is of LCP, then the receiving antenna must be of LCP. If it were of RCP, no signals would be produced at the

output of the receiving antenna. If it were linearly polarized, only half of the incoming power could be received and the other half would be lost, i.e. there would be a 3 dB loss in power.

If the incoming wave is linearly polarized and the receiving antenna is circularly polarized, is there a problem of polarization mismatch? The short answer is yes, depending on the combination scheme once the signal is received. Generally speaking, the power loss could be up to 3 dB due to polarization mismatch, which is much better than using a linearly polarized antenna with orthogonal polarization! In some applications, the polarization of the incoming signal is unknown or hard to predict. For example, if a linearly polarized wave is employed for satellite communications, its polarization will be changed when passing through the ionosphere due to Faraday rotation (which is caused by the free electrons in the ionosphere; the polarization rotation is determined by the electrons and the wave frequency) – thus, a circular polarization is usually employed. For radio communications in multipath environments (such as indoor), linear polarization is used – because the signal polarization becomes random after multiple reflections and could be said to be statistically uniform. The power loss due to polarization is about 3 dB. If circular polarization is employed, the power loss may be much more significant.

5.4.4 Radomes, Housings and Supporting Structures

The antenna is a transmitting and receiving device for a radio system and is normally made of metal. It is not an isolated structure but an integrated part of a radio system. Sometimes a radome is required in order to ensure its efficient and desired operation. The *radome* (radar dome) is defined as the structure which houses an antenna (or the entire radar, including the antenna) and is almost transparent to radio waves at the desired frequencies (normally above 1 GHz). It can be constructed in several shapes (spherical, geodesic, planar, etc.) depending upon the particular application. Various low-loss construction materials [3] are used, such as fiberglass, foam, FTFE and coated fabric or plastic. The main reasons for using such a radome or housing are

- to protect an antenna from the ravages of the environment: wind, snow, ice, rain, salt, sand, the sun (UV) and even lightning;
- to make the whole structure mechanically sound and viable;
- to conceal the antenna system from spying or the public view; and
- to keep nearby personnel away from being accidentally struck by a fast-moving antenna or affecting the antenna performance.

A few examples of antennas with radomes or housings are shown in Figure 5.56. Radomes are normally much larger than the antenna. Some housing structures are just slightly larger than the antenna and may not be called radomes, such as the mobile phone case, but they play the same role as radomes. Two of the most important questions asked in practice are

- How are these radomes and housing structures designed?
- What are the effects of the radome or housing on the antenna performance?

5.4.4.1 Design of Radomes and Housings

The antenna radome or housing design is a complex issue. Ideally, it should be incorporated into the antenna design; that is, it should be considered part of the antenna system. In practice,

Figure 5.56 Examples of antennas with various radomes/housings. (Reproduced by permission of Guidance Ltd, UK)

this is problematic. The radome or housing design may be conducted by another department or company. The basic requirement is that the radome or housing structure should be transparent to the operational frequencies of the antenna, which means that the reflection coefficient and loss must be small and the transmission coefficient must be large (as close to 1 as possible). From Section 3.3.1, we know that this requirement can be achieved when *the thickness of the structure is an integer multiple of half of the effective wavelength in the medium*. This thickness is actually a function of the incident angle as well as the frequency and dielectric properties (ε, μ, σ). The material is normally of low loss and low permittivity. Since the radome or housing has to meet both the electrical and mechanical requirements, the structure is often made from a multilayered material. The sandwich configuration is chosen: hard and reinforced materials are used to form the outer layers and a thick layer of low-permittivity material is placed at the center. This configuration provides the mechanically required robustness and at the same time offers the broad bandwidth and low loss for the RF signals.

5.4.4.2 Effects of Radomes and Housings

The effects of the radome and housing on the antenna include the following two aspects.

- From the field point of view, the existence of the radome or housing may affect the antenna aperture field distribution and hence the radiation pattern and directivity/gain. A slightly

reduced gain (<1 dB) and increased side lobes are expected in practice, depending on the quality of the radome/housing.

- From the circuit point of view, the load impedance to the antenna feed line may be changed, depending on how close the housing structure and the antenna are. The closer they are, the larger the effect. The effect may be reflected on the input impedance and hence the VSWR. A common observation is that the resonant frequency is shifted downwards (is lower) when the antenna housing is in place, since the wavelength is decreased inside a medium (by $\sqrt{\varepsilon_r}$).

5.4.4.3 Antenna Supporting Structure

Once an antenna is designed and made, it has to be installed somewhere. TV reception antennas are often erected on a roof and mobile base-station antennas are normally attached to a post or tower, as shown in Figure 5.57. Ideally, the antenna siting or supporting structure should have nothing to do with the antenna performance. However, the antenna is a radiating device and the radiation is defined by its radiation pattern. The antenna near-field distribution is more complex than that of the far field. Generally speaking, the supporting structure should not be in the radiation region and a dielectric material is preferred over a metal structure, since its interaction with the antenna is much smaller.

As the interaction between the antenna and the housing and supporting structure can only be analyzed case by case, there are no general formulas or equations for the prediction. The accurate estimate of this interaction should be conducted using computer simulation software, which could be slow and expensive. But sometimes it is worth paying such a cost.

Figure 5.57 Installed antennas and supporting structure

5.5 Summary

In this chapter we have studied many popular antennas, introduced all relevant theories and showed how to use these theories for antenna analysis and design. Wire-type antennas, aperture-type antennas and antenna arrays have been covered. These antennas have been dealt with from both the field point of view and the circuit point of view. Almost all the antennas introduced have been provided with design procedures or guidelines. Some practical considerations (including baluns and housings) have also been addressed.

References

[1] R. E. Collin, *Antennas and Radiowave Propagation*. McGraw-Hill, Inc., 1985.
[2] C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd edition, John Wiley & Sons, Inc., 1997.
[3] R. C. Johnson, *Antenna Engineering Handbook*, 3rd edition, McGraw Hill, Inc., 1993.
[4] J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd edition, McGraw Hill, Inc., 2002.
[5] S. Adachi and Y. Mushiake, 'Studies of large circular loop antenna', *Science Report, Research Institute of Tohoko University*, B.9.2, 1957, pp. 79–103.
[6] H. Morishita, K. Hirasawa and T. Nagao, 'Circularly polarised wire antenna with a dual rhombic loop,' *IEE Proc.-Microw. Antennas Propag*., **145**(3), 219–224, 1998.
[7] R. L. Li, V. Fusco and H. Nakano, 'Circularly polarized open-loop antenna,' *IEEE Trans. Antennas Propagat*., **51**(9), 2475–2477, 2003.
[8] M. Sumi, K. Hirasawa and S. Shi, 'Two rectangular loops fed in series for broadband circular polarization and impedance matching,' *IEEE Trans. Antennas Propagat*., **52**(2), 551–554, 2004.
[9] G. Kandonian and W. Sichak, 'Wide frequency range tuned helical antennas and circuits,' *IRE Conv Rec*. Part 2, 1953, pp. 42–47.
[10] H. Yagi, 'Beam transmission of ultra-short waves,' *Proc IRE*, **16**, 715, 1928.
[11] P. P. Viezbicke, 'Yagi Antenna Design,' *NBS Technical Note* 688, December, 1968.
[12] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd edition, John Wiley & Sons, Inc., 1998.
[13] C. A. Chen and D. K. Cheng, 'Optimum spacings for Yagi–Uda arrays,' *IEEE Trans. on Antenna and Propagation*, **AP-21**, 615–623, 1973.
[14] C. A. Chen and D. K. Cheng, 'Optimum element lengths for Yagi–Uda arrays,' *IEEE Trans. on Antenna and Propagation*, **AP-23**, 8–15, 1975.
[15] S. Baskar, A. Alphones, P.N. Suganthan and J.J. Liang, 'Design of Yagi–Uda antennas using comprehensive learning particle swarm optimization,' *IEE Proc Microwave, Antennas and Propagation*, **152**(5), 340–346, 2005.
[16] S. Lim and H. Ling, 'Design of a closely spaced, folded Yagi antenna', *IEEE Antennas Wireless. Propag. Lett.*, **5**, 302–305, 2006.
[17] Z. Bayraktar, P. L. Werner and D. H. Werner, 'The design of miniature three-element stochastic Yagi–Uda arrays using particle swarm optimization,' *IEEE Antennas Wireless. Propag. Lett.*, **5**, 22–26, 2006.
[18] R. L. Carrel, 'Analysis and design of log-periodic dipole antenna,' PhD Thesis, University of Illinois, 1961.
[19] P. C. Butson and G. T. Thompson, 'A note on the calculation of the gain of log-periodic dipole antennas,' *IEEE Trans on Antennas and Propagation*, **AP-24**, 105–106, 1976.
[20] J. Aurand, 'Pyramidal horns, Part 2: Design of horns for any desired gain and aperture phase error,' *IEEE Antennas & Propagation Society Newsletter*, **31**, 25–27, 1989.
[21] W. Love (Ed.), *Reflector Antennas*, IEEE Press, New York, 1978.
[22] J. R. James and P. Hall, *Handbook of Microstrip Antennas*, IEE, London, 1988.
[23] D. R. Jackson and N. G. Alexopoulos, 'Simple approximate formulas for input resistance, bandwidth and efficiency of a resonant rectangular patch,' *IEEE Trans AP*, 407–410, 1991.
[24] R. Garg, Prakash Bhartia and Inder Bahl, *Microstrip Antenna Design Handbook*, Artech House Publishers, 2000.
[25] Z. N. Chen and M. Y. W. Chia, *Broadband Planar Antennas: Design and Applications*. John Wiley & Sons, Ltd, 2006.

[26] K. R. Carver and J. W. Mink, 'Microstrip antenna technology,' *IEEE Trans on Antenna and Propagation*, **29**, 2–24, 1981.
[27] R. S. Elliott, *Antenna Theory and Design*, Prentice Hall, Englewood Cliffs, 1981.
[28] W. W. Hansen and J. R. Woodyard, 'A new principle in directional antenna design,' *Proc. IRE*, **26**, 333–345, 1938.
[29] C. L. Dolph, 'A current distribution for broadside arrays which optimizes the relationship between beamwidth and side-lobe level,' *Proc. IRE and Waves and Electrons*, June, 1946.
[30] B. J. Forman, 'Directivity of scannable planar arrays,' *IEEE Trans. Antennas and Propagation*, **20**, 245–252, 1972.
[31] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill, New York, 1961.

Problems

Q5.1 From an antenna point of view, a two-wire transmission line can be considered as two wire antennas. Explain why this type of transmission is only suitable for low-frequency, not high-frequency, applications.

Q5.2 Dipole antennas are popular, can you justify why the half-wavelength dipole is the most popular dipole?

Q5.3 Compare the half-wavelength dipole, quarter-wavelength monopole and half-wavelength monopole in terms of the input impedance, directivity and radiation pattern.

Q5.4 How can the bandwidth of the half-wavelength dipole be broadened?

Q5.5 Compare a half-wavelength dipole, a half-wavelength slot and a half-wavelength loop in terms of the input impedance, gain and radiation pattern.

Q5.6 A short dipole with a length of 3 cm and a diameter of 3 mm is made of copper wire for 433 MHz (an ISM band) applications. Find its input impedance, radiation resistance and radiation efficiency. If it is to be connected to a 50 Ω coaxial cable, find the reflection coefficient, return loss and VSWR.

Q5.7 Explain the concept of baluns and design a balun for Q5.6 if applicable.

Q5.8 Design a circularly polarized helix antenna of an end-fire radiation pattern with a directivity of 10 dBi. Find out its input impedance, *HPBW*, *AR* and radiation pattern.

Q5.9 Explain the design principle of the Yagi–Uda antenna and estimate the maximum directivity of a Yagi–Uda antenna with ten elements.

Q5.10 Discuss how the directivity is linked to the number of elements, N, scaling factor and apex angle of a log-periodic dipole antenna (LPDA). If the directivity of a six-element LPDA is 11 dBi, what is the achievable bandwidth?

Q5.11 Design a log-periodic dipole antenna to cover all UHF TV channels, which is from 470 MHz for channel 14 to 890 MHz for channel 83 and each channel has a bandwidth of 6 MHz. The desired directivity is 12 dBi.

Q5.12 Design a gain horn antenna with a directivity of 15 dBi at 10 GHz. A WR-90 waveguide is used to feed the horn.

Q5.13 The Sky-digital satellite receiving antenna is an offset parabolic reflector antenna with a dish diameter of about 40 cm. If the aperture efficiency factor is about 70%, estimate the directivity of this antenna. The central frequency is 12 GHz.

Q5.14 A circular parabolic reflector has $F/2a = 0.5$. The field pattern of the feed antenna is $E(\theta) = \cos^2\theta$, $\theta < \pi/2$. Find the edge taper, spillover efficiency and aperture efficiency.

Q5.15 RT/Duroid 6010 substrate ($\varepsilon_r = 10.2$ and $d = 1.58$ mm) is to be used to make a resonant rectangular patch antenna of linear polarization.

a) Design such an antenna to work at 2.45 GHz for Bluetooth applications.
b) Estimate its directivity.
c) If it is to be connected to a 50 ohm microstrip using the same PCB board, design the feed to this antenna.
d) Find the fractional bandwidth for *VSWR* < 2.

Q5.16 Two half-wavelength dipoles are in parallel and separated by $\lambda/4$. Find the input impedance for each antenna and then obtain and plot the radiation pattern for the following cases:
a) they are fed in phase;
b) they are fed 90 degrees out of phase;
c) they are fed 180 degrees out of phase.

Q5.17 Four half-wavelength dipoles are employed to form a linear array. Find the maximum directivity obtainable for this array and plot its radiation pattern.

Q5.18 Four $4/\lambda$ monopoles with a large ground plane are used to form a linear array. The spacing between elements is 1λ, all monopoles are to be fed in phase from a 1 GHz transmitter via a 50 Ω microstrip line.
a) Design a feed system for this array (mutual coupling is ignored).
b) Find its radiation pattern. Is this a broadside or end-fire array?
c) Find its directivity.

Q5.19 Explain the principle of a phase scanned array.

Q5.20 What is the Hansen–Woodyard end-fire array?

Q5.21 Design a 1.5 GHz circularly polarized antenna that has an omnidirectional radiation pattern in the H-plane.

Q5.22 The bandwidth is an important consideration for any RF device. An antenna with an impedance of 300 Ω is matched to a 50 Ω feed line with a quarter-wavelength transformer at the design frequency 1 GHz. Find the bandwidth for *VSWR* < 2 if the antenna impedance is a constant, and then comment on your results if the antenna impedance is not a constant (the real case).

Q5.23 Double $\lambda/4$ transformers can improve the bandwidth of the matching network. Redo Q5.22 using a double $\lambda/4$ transformer.

6

Computer-Aided Antenna Design and Analysis

In the previous chapter we introduced and analyzed some of the most popular antennas, and discussed what the important considerations are from the design point of view. You should now be able to design some basic standard antennas (such as dipoles) without resorting to a computer. However, antenna design is a very challenging subject since there are usually many variables involved, even for a given type of antenna. Due to significant advances in computer hardware and software over the past two decades, it has now become standard practice to employ software to aid antenna design – just as in any other industry. The objective of this chapter is therefore to give a brief review of antenna software development, introduce the basic theory behind computer simulation tools and demonstrate how to use industrial standard software to analyze and design antennas.

6.1 Introduction

As the antenna equation (4.3) indicates, the prediction of antenna performance is a very complex issue and the analytical approach is only suitable for antennas of simple geometry. In the past, engineers designed antennas by experience and a certain amount of 'black magic'. A good understanding of antennas and impedance-matching theory and techniques, and the ability to accomplish a good match over the desired frequency range were required. After the prospective antenna was designed and built, there always remained the question of whether the antenna had met the specifications. Thus, an antenna measurement using a test range had to be conducted in order to verify that all design criteria were met. Next, if all went well on the testing, the prospective antenna would be transferred to the production department. Finally, the production antennas would be tested to verify that the design could be duplicated. However, if the measured results were not good enough, some tuning and modifications had to be made and further tests were required. This process normally included minding, tinkering, cut-and-try and gluing and screwing parts together until the antenna met the desired specifications. Needless to say, this was a very expensive and time-consuming procedure.

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

Those days are now gone forever. We live in a competitive market where time and performance are critical. Guesswork and long design cycles are out of the question. New state-of-the-art antennas are required to meet all the design criteria for sophisticated applications that may involve large quantities at competitive cost. Various commercial antenna design software packages have been developed with very good accuracy and multi-functions and these are widely used in the industry. The ability to use computer-aided design (CAD) tools has become an essential requirement for a good antenna engineer.

Now let us look back to see how the antenna design software has evolved. In the mid-1960s, mainframe computers were finally fast enough to conduct antenna modeling. However, it still required an astute antenna design engineer with a good analytic mathematical background to write the equations and computer programs to model antennas. In the 1970s, the Numerical Electromagnetic Code (NEC) program was developed using FORTRAN (a computer computational program) by the US government [1]. This program was accurate and suitable for wire-type antennas but difficult to use and required a large mainframe computer. Furthermore, its use was restricted by the government. This all changed in the mid-1980s when personal computers (PCs) became widely available. Soon there were several antenna-modeling programs available. The primary one was MININEC, a smaller program based on NEC that ran fast on a PC [2 and 3]. Also, a PC version of NEC became available. As PCs increased in speed, so did the antenna-modeling programs. It was not unusual to wait for hours or more for the analysis of an antenna. Nowadays, PCs with powerful processors can model the same antenna in just a few seconds. In the 1990s, a large number of EM simulation tools and antenna design software packages, based on various methods, were developed and appeared on the market. They are suitable not only for wire-type antennas but also aperture-type antennas. Many of these packages have been updated every one or two years, with new functions and improved algorithms added. For example, parametric optimization is now available. These programs will perform the optimization for you. All you need to do is to input a reasonable design along with a properly weighted trade-off or figure of merit (FoM). The FoM weighs the relative importance of each antenna parameter such as the gain, pattern and impedance over a set of input frequencies and an optimized design will then be produced by the computer. These powerful modeling programs allow the antenna designer to use keystrokes on a PC keyboard instead of getting their hands dirty from hours of cutting and testing the actual antenna. Furthermore, the performance goals can be tested before any actual antenna is constructed. You can now cut-and-try an antenna design for maximum gain, front-to-back ratio, impedance or a combination of the three by changing element lengths, diameters and spacings using only a keyboard. This significantly shortens the design to completion of production process for antennas.

The above may give the impression that antenna design is no longer a ‘black art’. While this may be true, it still requires a lot of knowledge and understanding of the interrelated parameters in the antenna design, FoM and the software itself. In fact, as the antenna designs of today get more complicated, so does the work of the antenna designer. The antenna-modeling and optimization programs available today do not completely compensate for all variables. The output data file must be properly interpreted by the user. Sometimes the tolerances of the element dimensions may be difficult to realize. Element correction factors may still be required to compensate for the mechanical mounting of the antenna elements. Sometimes the simulation results may not be accurate enough for practical reasons (such as limited computer memory or poor convergence). This is where the antenna designer comes in to play. One must be able to realize which designs can and cannot be built, in addition to any compensation that may be

required. After all, computers are just computers and their outputs are just numbers, they will never replace innovative engineers in analyzing and designing antennas. Modeling and design software do not design an antenna for you. They only predict how the inputted antenna design should perform, and leave it up to the ingenuity of the antenna designer to 'tweak' or optimize the mechanical dimensions until the desired antenna performance is attained.

In summary, to become a modern professional antenna designer, one has to have a comprehensive understanding of antennas, a good grasp of mechanical engineering principles and excellent skills in using computer-modeling and optimization software to aid the design.

6.2 Computational Electromagnetics for Antennas

Computational electromagnetics (CEM) is a unique subject of interest to all electromagnetics engineers and researchers. It has a very wide range of applications in RF engineering, EMC, radar, wireless communications, electrical and electronic engineering, and extends to areas such as biomedical engineering – antennas are just one of the areas. Many methods have been developed over the years; some were first introduced for other applications. Generally speaking, CEM can be divided into numerical methods and high-frequency methods, as shown in Figure 6.1. *High-frequency methods are suitable for structures much larger than the wavelength, while numerical methods are more suitable for smaller structures.* There is no specific boundary between these two groups. As a rule of thumb, 20λ is usually used as the upper limit for numerical methods, which is really determined by the computation power (memory size and computational time).

Numerical methods can be subdivided into frequency domain methods (such as the method of moments and the finite element method) and time domain methods (e.g. the finite-difference time domain method and the transmission line modeling method). High-frequency methods can be subdivided into field-based methods (the geometrical optics method) and current-based methods (the physical optics method). Although cost-effective electromagnetic software is readily available and it is now not necessary to write your own software from scratch, a good understanding of the principles on which the software is based is necessary in order to set the relevant parameters properly and avoid the misuse and misinterpretation of the results.

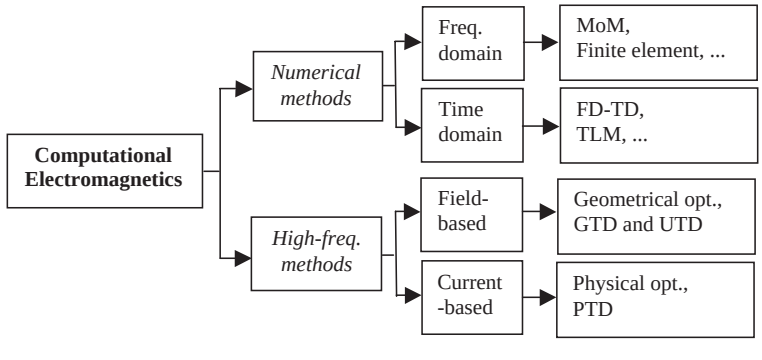


Figure 6.1 Classification of computational electromagnetic methods

In this section we are going to first use the method of moments as an example to see how a numerical method can be employed to model and analyze an antenna, then we will introduce some other methods. A general comparison of these methods will be made in Section 6.2.5.

6.2.1 Method of Moments (MoM)

The *method of moments* (MoM, or moment method) was first introduced in mathematics. The basic idea is to transform an integral or differential equation into a set of simultaneous linear algebraic equations (or matrix equation) which may then be solved by numerical techniques [4, 5, 6]. It was first applied to electromagnetic problems in the 1960s. Roger Harrington was the person who made the most significant contribution to this area [5, 6]. His book, *Field Computation by Moment Methods* [6], was the first book to explore the computation of electromagnetic fields by the method of moments – the most popular method to date for the numerical solution of electromagnetic field problems. It presents a unified approach to MoM by employing the concepts of linear spaces and functional analysis. Written especially for those who have a minimal amount of experience in electromagnetic theory, theoretical and mathematical concepts are illustrated by examples that prepare readers with the skills they need to apply the method of moments to new, engineering-related problems. In this subsection we are going to introduce the concept of MoM and apply it to model and analyze a dipole antenna.

6.2.1.1 Introduction to MoM

Let us deal with a general case: a linear equation

$$L(F) = g \tag{6.1}$$

where L is a known linear operator, g is a known excitation function and F is the unknown function to be determined. In physics, L means the system transfer function and g represents the source. The objective is to determine F once L and g are specified.

First, the function F is expanded using a series of known *basis functions* (or *expansion functions*) $f_1, f_2, f_3, \ldots$ in the domain of operator L, we have

$$F = \sum_n I_n f_n \tag{6.2}$$

where I_n are unknown complex coefficients to be determined and $n = 1, 2, \ldots N$. N should be infinite in theory but is a limited number in practice.

We can replace F in Equation (6.1) by Equation (6.2) and use the linearity of L to give

$$\sum_n I_n L(f_n) = g \tag{6.3}$$

The problem becomes how to determine these unknown coefficients I_n.

Secondly, a set of *weighting functions* (or *testing functions*) $W_1, W_2, W_3, \ldots$ in the domain of L is chosen and then the inner product is formed:

$$\sum_n I_n \langle W_m,\ L(f_n)\rangle = \langle W_m,\ g\rangle \tag{6.4}$$

where $m = 1, 2, \ldots M$. Again, M should be infinite in theory but is a finite number in practice and a typical, but not unique, *inner product* is defined as

$$\langle x(z), y(z)\rangle = \langle y(z), x(z)\rangle = \int_L x(z)y(z)dz \tag{6.5}$$

Thirdly, the inner product is performed on Equation (6.4) for $m = 1$ to M to give the matrix equation:

$$\begin{bmatrix} < W_1, L(f_1) > & < W_1, L(f_2) > & \ldots & < W_1, L(f_N) > \\ < W_2, L(f_1) > & < W_2, L(f_2) > & & < W_2, L(f_N) > \\ \quad \ldots & & & \\ < W_M, L(f_1) > & \ldots & & < W_M, L(f_N) > \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ . \\ I_N \end{bmatrix} = \begin{bmatrix} < W_1, g > \\ < W_2, g > \\ . \\ < W_M, g > \end{bmatrix} \tag{6.6}$$

or, in more compact form:

$$[Z_{mn}][I_n] = [V_m] \tag{6.7}$$

where

$$Z_{mn} =< W_m, L(f_n) > \tag{6.8}$$

and

$$V_m =< W_m, g > \tag{6.9}$$

They are now readily obtained and the unknown coefficients can be yielded as

$$[I_n] = [Z_{mn}]^{-1}[V_m] \tag{6.10}$$

The unknown function F can therefore be obtained approximately using Equations (6.2) and (6.10).

The selection of the basis functions and weighting functions is a key to obtaining accurate solutions efficiently and successfully. In general, the basis functions should have the ability to accurately represent and resemble the anticipated unknown function. There are many possible basis/weighting function sets, but only a limited number are used in practice. The elements of these functions must be linearly independent, so that the N equations in Equation (6.6) will also be linearly independent and give a unique solution. In addition, they should be chosen to

minimize the computations required to evaluate the inner product. If both the basis functions and weighting functions are the same, then this special procedure is known as *Galerkin's method*. We are going to use the example below to show how to implement the MoM in practice.

Example 6.1: MoM. Solve the following differential equation using the MoM:

$$\frac{d^2F(z)}{dz^2} = 1 + 2z^2 \tag{6.11}$$

for $0 \leq z \leq 1$with the boundary conditions: $F(0) = F(1) = 0$.

Solution:
This is a simple boundary problem. The transfer function is

$$L = \frac{d^2}{dz^2}$$

and the source is

$$g(z) = 1 + 2z^2$$

It is not difficult to obtain the exact solution of this differential equation as

$$F(z) = -\frac{2}{3}z + \frac{1}{2}z^2 + \frac{1}{6}z^4 \tag{6.12}$$

There are many basis and weighting functions which may be suitable for this problem. The selection of these functions is crucial. Here we are going to use Galerkin's method to solve this problem.

According to the format of the source $g(z)$, it is natural to choose the basis function as $f_n(z) = z^n$. However, you will find in the end that the boundary condition $F(1) = 0$ cannot be met. Taking the boundary conditions into account, we can choose the basis functions and weighting functions as

$$f_n(z) = W_n(z) = z - z^{n+1} \tag{6.13}$$

and the inner product is defined as

$$< f_m(z),\ f_n(z) >= \int_0^1 f_m(z)\, f_n(z) dz \tag{6.14}$$

Using Equations (6.8) and (6.9), we can obtain

$$\begin{aligned} Z_{mn} &= < W_m, L(f_n) >= \int_0^1 (z - z^{m+1}) \frac{d^2}{dz^2}(z - z^{n+1}) dz = -\frac{mn}{m+n+1} \\ V_m &= < W_m, g >= \int_0^1 (z - z^{m+1})(1 + 2z^2) dz = \frac{m^2 + 3m}{(m+2)(m+4)} \end{aligned} \tag{6.15}$$

For $M = N = 1$, we have $Z_{11} = -1/3$ and $V_1 = 4/15$. We now use Equation (6.10) to give

$$I_1 = -\frac{4}{5}, \text{ and the solution: } F_{N=1} = \sum_n I_n f_n = -\frac{4}{5}(z - z^2)$$

As shown in Figure 6.2, the result obtained by the MoM (dashed line) is close to the exact solution (solid line). To increase the accuracy, we can increase N.

For $M = N = 2$, we have

$$\begin{bmatrix} -\frac{1}{3} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{4}{5} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} \frac{4}{15} \\ \frac{5}{12} \end{bmatrix}$$

Thus, we can easily use a computer program (such as Matlab [7]) to solve the matrix equation to obtain

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} -\frac{3}{10} \\ -\frac{1}{3} \end{bmatrix}$$

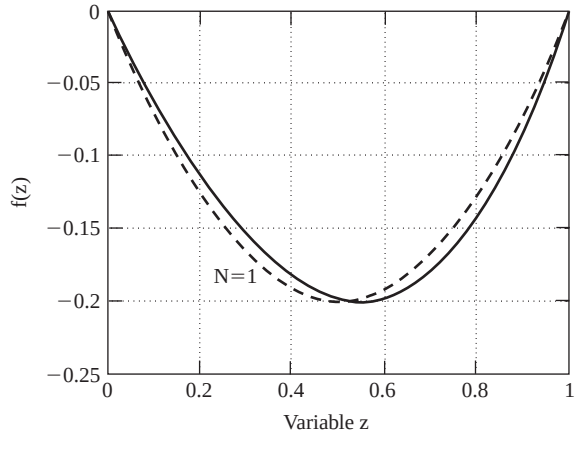


Figure 6.2 Comparison of MoM result for $N = 1$ (dashed line) with the exact solution (solid line)

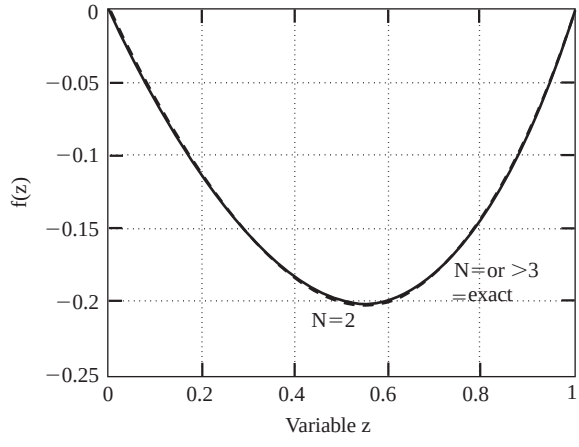


Figure 6.3 Comparison of MoM result for $N = 2$ (dashed line) with the exact solution (solid line)

and

$$F_{N=2} = \sum_n I_n f_n = -\frac{19}{30}z + \frac{3}{10}z^2 + \frac{1}{3}z^3$$

This is very close to the exact solution, as shown in Figure 6.3. If we increase M and N further to, say, 3, then we can yield

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ -\frac{1}{6} \end{bmatrix}$$

thus

$$F_{N=3} = -\frac{2}{3}z + \frac{1}{2}z^2 + \frac{1}{6}z^4$$

This is now the same as the exact solution given by Equation (6.12). It is not difficult to verify that, when $N > 3$, the solutions are unchanged, i.e. the same as Equation (6.12). From Figures 6.2 and 6.3 we can see that the accuracy of the numerical solutions increases with N, and this means that the results converge. It is very important to have a convergent solution. If the result does not converge, it normally indicates that the basis and weighting functions have not been properly chosen and are not suitable for the problem.

This example has successfully demonstrated the usefulness of the MoM in solving linear equations. Through this exercise we have seen that a differential equation has been converted to a matrix equation, which is easy to solve and the results are very accurate.

Next, we are going to see how to use the MoM to analyze and model a dipole antenna.

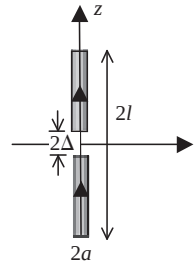


Figure 6.4 A dipole of length $2l$, diameter $2a$ and gap 2Δ

6.2.1.2 Analysis of a Dipole Antenna Using the MoM

The MoM has been widely used to solve wire-type antenna problems. Here we are going to take a dipole of length $2l$ and diameter $2a$ as an example, shown in Figure 6.4, to illustrate the application of the MoM in antennas.

The total electric field at any point in the space can be expressed as the sum of the incident field $\boldsymbol{E^i}$ and scattered/radiated field $\boldsymbol{E^s}$:

$$\boldsymbol{E^t}(\boldsymbol{r}) = \boldsymbol{E^i}(\boldsymbol{r}) + \boldsymbol{E^s}(\boldsymbol{r}) \tag{6.16}$$

Since the dipole is made of a perfect conductor (a good conductor in reality), the total tangential electric field must be zero on the surface of the antenna, that is

$$E_z^i(r) + E_z^s(r) = 0; \text{on the antenna} \tag{6.17}$$

For a receiving antenna or scatterer, the incident field is usually a plane wave. For a transmitting antenna, the incident field can be seen as the excitation of the antenna and there are two source models: the delta gap source and the magnetic frill generator [8, 9]. The most used source model is the delta gap model, which is also chosen for our investigation. In Figure 6.4, a voltage source of $2V_0$ (from $+V_0$ to $-V_0$) is applied across the feeding gap 2Δ, the incident field can therefore be expressed as

$$E_z^i(r) = \begin{cases} V_0/\Delta; \text{ for } |z| < \Delta \\ 0; \text{ for } \Delta < |z| < l \end{cases} ; \text{on the antenna} \tag{6.18}$$

The general radiated field can be represented by the antenna equation (4.3) as

$$\boldsymbol{E}^s(\boldsymbol{r}) = \left(-j\omega\mu + \frac{\nabla\nabla\bullet}{j\omega\varepsilon}\right) \int_V \boldsymbol{J}(\boldsymbol{r}') \frac{e^{-j\beta\cdot|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi\,|\boldsymbol{r}-\boldsymbol{r}'|} dv' \tag{6.19}$$

On the surface of the antenna, the current density can be replaced by a line-source current $I(z)$, we have

$$\int_{-l}^{l} I(z')K(z,z')dz' = -E_z^i(z) \tag{6.20}$$

where the kernel is given by

$$K(z,z') = \frac{1}{j4\pi\omega\varepsilon}\left(\frac{\partial^2}{\partial z^2} + \beta^2\right)\frac{e^{-j\beta r}}{r} \tag{6.21}$$

Equation (6.20) is known as *Pocklington's integral equation* and has been used extensively for dipole antennas. The problem has now become how to determine the current $I(z)$ by solving this integral equation.

Using the MoM, we need to choose a set of basis functions. The operator in this case is an integral with a complex kernel. Although we can select the entire domain basis functions as in Example 6.1, we prefer to use sub-domain functions, which should make the inner product computation much simpler and faster.

The antenna can be divided equally into N segments, as shown in Figure 6.5. The unit pulse function is defined as

$$P_n(z) = \begin{cases} 1; & \text{for } -l + \dfrac{2l}{N}(n-1) < z \le -l + \dfrac{2l}{N}n \\ 0; & \text{otherwise} \end{cases} \tag{6.22}$$

and is chosen for the basis functions, thus the current on the dipole is approximated as

$$I(z) \approx \sum_{n=1}^{N} I_n P_n(z) \tag{6.23}$$

This expansion in terms of pulse functions is a staircase approximation. The current is constant on each segment. The larger N, the more accurate the solution should be.

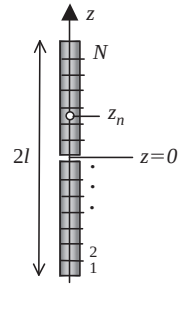


Figure 6.5 A dipole is equally divided into N segments

We can use Galerkin's method again, which can produce a reasonable solution. A better weighting function for this case is the Dirac delta function $\delta(z)$, which is

$$W_m(z) = \delta(z - z_m) \tag{6.24}$$

where z_m is the z-coordinate of the center of the segment m, mathematically it is

$$z_m = -l + \frac{2l}{N}(m - 0.5); \text{ and } m = 1,\ 2, \ldots, N \tag{6.25}$$

A very useful feature of the Dirac delta function is

$$\int f(z)\delta(z - z_0)dz = f(z_0) \tag{6.26}$$

This approach is called the *pulse-expansion and point-matching* MoM. The integral equation is enforced at N points along the antenna axis. The inner product can be defined as

$$< f_m(z),\ f_n(z) >= \int_{-l}^{l} f_m(z)\, f_n(z)dz \tag{6.27}$$

Thus, the element

$$\begin{aligned} Z_{mn} =< W_m, L(f_n) > &= \int_{-l}^{l} [\delta(z - z_m) \int_{-l}^{l} P_n(z')K(z, z')dz']dz \\ &= \int_{-l+2l(n-1)/N}^{-l+2l(n)/N} K(z_m, z')dz' \end{aligned} \tag{6.28}$$

and

$$V_m =< W_m, g >= -\int_{-l}^{l} \delta(z - z_m)E_z^i(z)dz = -E_z^i(z_m) \tag{6.29}$$

We can let N be an odd number and choose the central segment as the feed to simplify the computation.

It should be pointed out that if the matching point and the source point are at the same place, the distance r will be zero and the kernel will be infinite – this is a problem. A reasonable way to avoid this singularity is to let the current source be at the center of the wire and the matching point on the antenna surface, that is

$$R_m = r\left(z_m, z'\right) = \sqrt{a^2 + (z_m - z')^2} \tag{6.30}$$

Equation (6.28) can be written in a more computationally friendly form as [10, 11]

$$Z_{mn} = \int_{-l+2l(n-1)/N}^{-l+2l(n)/N} \frac{e^{-j\beta R_m}}{j4\pi\omega\varepsilon R_m^5}\left[(1+j\beta R_m)\left(2R_m^2-3a^2\right)+(\beta a R_m)^2\right]dz' \tag{6.31}$$

When the segment is small enough, this expression can be further simplified to

$$Z_{mn} \approx \frac{e^{-j\beta R_{mn}}}{j4\pi\omega\varepsilon R_{mn}^5}\left[(1+j\beta R_{mn})\left(2R_{mn}^2-3a^2\right)+(\beta a R_{mn})^2\right]\frac{2l}{N} \tag{6.32}$$

where

$$R_{mn} = r(z_m, z_n) = \sqrt{a^2+(z_m-z_n)^2} \tag{6.33}$$

Now Pocklington's integral equation (6.20) is transformed into the following matrix equation:

$$[Z_{mn}][I_n] = [V_m] \tag{6.34}$$

All elements of Z_{mn} and V_m can be computed easily and the unknown coefficients I_n can now be obtained without difficulty. Hence, the current distribution along the antenna can be found using Equation (6.23) and is illustrated by Figure 6.6.

Other important antenna parameters can also be obtained. For example, the input impedance is given by

$$Z_a = \frac{2V_0}{I_{N/2}} \tag{6.35}$$

and the total radiated field (and hence the radiation pattern) can be calculated using

$$\boldsymbol{E}^t(\boldsymbol{r}) = \boldsymbol{E}^i(\boldsymbol{r}) + \boldsymbol{E}^s(\boldsymbol{r}) = \boldsymbol{E}^i(\boldsymbol{r}) + \int_{-l}^{l}\sum_{n=1}^{N} I_n P_n(z')\boldsymbol{K}(\boldsymbol{r}, z')dz' \tag{6.36}$$

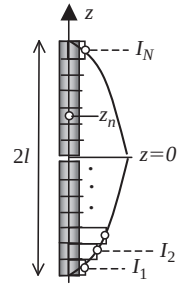


Figure 6.6 Current distribution along a dipole

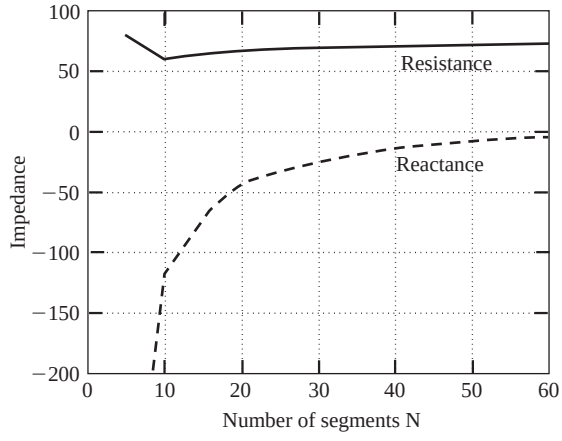


Figure 6.7 Input impedance convergence of the point-matching approach

For a half-wavelength dipole, the convergence of the calculated impedance as a function of the number of segments N is shown in Figure 6.7. When N is small, the impedance is not stable. When $N > 20$, the resistance approaches 80 ohms and the reactance is close to 0 ohms, which are comparable with measured results.

Using this program, we can produce a lot of results for different antennas, without actually making any antennas. From Figure 6.8, which is also shown in Table 5.1, we can clearly see how the current distribution varies with the antenna length in wavelengths and how the current distribution determines the radiation pattern.

6.2.1.3 Discussion and Conclusions

In this subsection we have introduced the MoM and demonstrated how to use it to solve a linear equation and obtain the important parameters of a dipole antenna. It is important

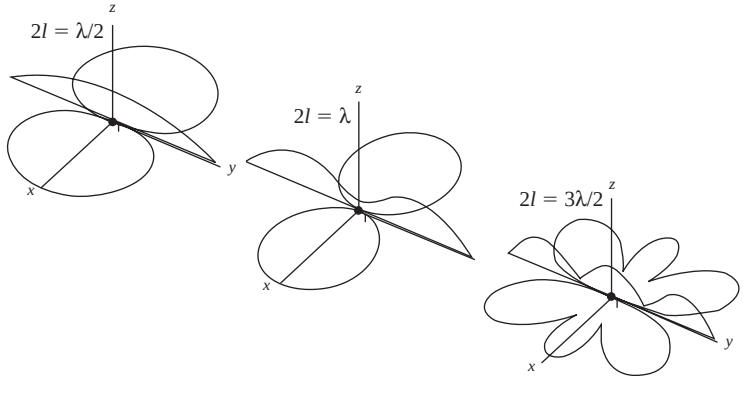


Figure 6.8 Current distributions and radiation patterns of three dipoles with lengths λ/2, λ and 3λ/2

to note:

- the selection of the basis and weighting functions is crucial, this may affect the complexity and time of computation as well as the accuracy of the results;
- if the result converges, then the more segments, the more accurate it will be (but with an increased memory requirement and computation time!);
- if the result does not converge, it usually means that the basis and weighting functions are not suitable;
- using the point-matching technique, the segment length should be smaller than $\lambda/10$, it is usually about $\lambda/20$, which is also generally required by other simulation methods;
- there is a trade-off between the accuracy and computation requirements (memory and time) – this is a conclusion that is applicable to all numerical methods.

Although we have only applied the MoM to a dipole antenna, the method is general and can be easily applied to other wire-type antennas. It can even be used to solve aperture-type antenna problems. Another beauty of it is that it can be combined with high-frequency methods to deal with electrically large problems. Some very good commercial software packages are already available on the market and will be discussed later in this chapter.

6.2.2 Finite Element Method (FEM)

The *Finite Element Method (FEM)* was originally introduced for solving complex elasticity, structural analysis problems in civil engineering and aeronautical engineering. Its development can be traced back to the work of Alexander Hrennikoff [12]. The method was often based on an energy principle, e.g. the virtual work principle or the minimum total potential energy principle, which provides a general, intuitive and physical basis that has a great appeal to structural engineers. A rigorous mathematical foundation was provided by Strang and Fix in their book *An Analysis of the Finite Element Method* published in 1973 [13], and the method has since been generalized into a branch of applied mathematics for numerical modeling of physical systems in a wide variety of engineering disciplines, e.g. electromagnetics and fluid dynamics. It has been used for finding approximate solutions of partial differential equations as well as integral equations, and is particularly suitable for problems involving irregular boundaries and nonhomogenous material properties.

The FEM may be implemented in the following four steps:

1. Discretization of the solution region into elements (usually triangular in shape).
2. Generation of equations for the fields or potentials at each element.
3. Integration or assembly of all elements.
4. Solution of the resulting system of equations.

Just like the MoM, the problem is also converted into a matrix equation at the end. The FEM has been employed by a large number of commercial EM simulation packages and has become one of the most popular and established numerical techniques in engineering. An example of how the region of a loop antenna is discretized into FEM elements is shown in Figure 6.9. The obtained current distribution on the loop is also shown in the figure. In Section 6.3 we are going to use FEM-based software to simulate and design antennas. For more details about this technique, refer to [14, 15].

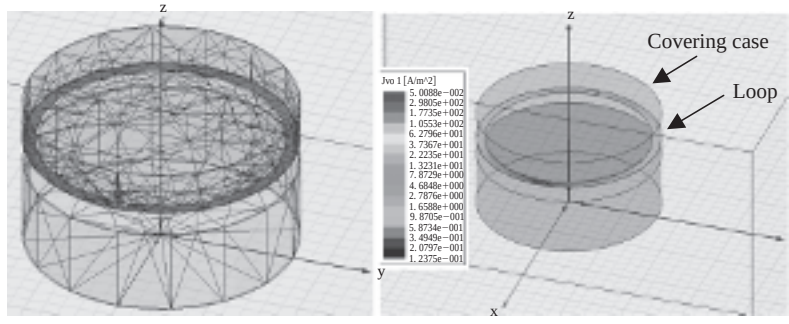


Figure 6.9 FEM simulation of a loop antenna: discretization and current distribution. In this case, the loop is placed in a device (e.g. a watch) for wireless communications, the effects of the covering case on the loop can be estimated

6.2.3 Finite-Difference Time Domain (FDTD) Method

The *Finite-Difference Time Domain* (FDTD) method belongs in the general class of grid-based differential time domain numerical modeling methods. It was introduced by Yee in 1966 [16]. Since then significant developments have been made in improving, implementing and spreading this method [17, 18, 19]. The boundary conditions were one of the major problems of this method.

The problem domain is discretized into many cells (usually square or rectangular), known as Yee cells/lattices. A typical one is shown in Figure 6.10. The time-dependent partial differential Maxwell equations given in Equation (1.29) are discretized using central-difference approximations to the space and time partial derivatives. The resulting finite-difference equations are solved in a leapfrog manner: the electric field vector components in a volume of space are solved at a given instant in time, then the magnetic field vector components in the same spatial volume are solved at the next instant in time. This process is repeated over and over again until the desired transient or steady-state electromagnetic field behavior is fully

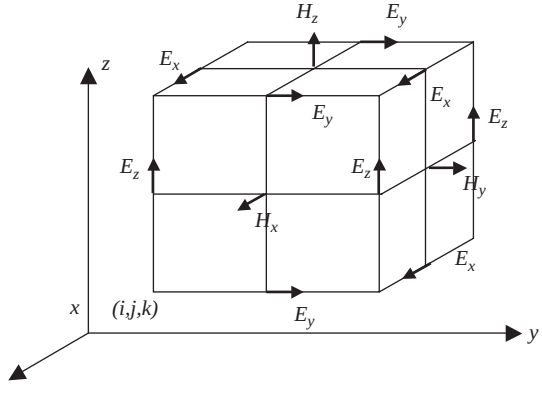


Figure 6.10 FDTD (Yee) cell

evolved. This scheme has proven to be very robust and remains at the core of many current FDTD software constructs. Furthermore, Yee proposed a leapfrog scheme for marching in time where, in the ***E***-field and ***H***-field, updates are staggered so that ***E***-field updates are conducted midway during each time step between successive ***H***-field updates, and conversely. On the plus side, this explicit time-stepping scheme avoids the need to solve simultaneous equations, and furthermore yields dissipation-free numerical wave propagation. On the minus side, this scheme mandates an upper bound on the time step to ensure numerical stability. As a result, certain classes of simulation can require many thousands (or more) of time steps for completion.

The accuracy of the computed results is basically determined by the cell size and time steps, which is simpler and easier than making the right choices on basis and weighting functions in the MoM. The cell size is determined by the highest frequency of interest. The biggest dimensions of the cell should normally be smaller than $\lambda_{\min}/20$, and the time step size $\Delta t \leq 1/c\sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2}}$. Thus, this is limited by the computational power as for any other numerical method. One of the advantages of this method is that material properties $(\varepsilon, \mu, \sigma)$ can be accommodated easily into the computational scheme, in just the way they appear in Maxwell's equations.

Since it is a time domain method, the method is extremely suitable for solving wideband problems, but narrowband problems may take a long time to converge. Today the FDTD has become the most popular time domain method in computational electromagnetics. A number of well-known computer simulation tools were developed based on this technique.

6.2.4 Transmission Line Modeling (TLM) Method

The *Transmission Line Modeling* (TLM) method is another well-known time domain modeling method, which was originally introduced by Professor P. B. Johns at the University of Nottingham, UK [20]. The main features of this method are its simplicity and the use of the transmission line model, which means that lumped elements are used. The method has evolved and been developed over the years [21, 22, 23]. The simulation domain is divided into many nodes; a typical one is shown in Figure 6.11. All fields (***E*** and ***H***) and material properties are represented by transmission line elements (such as V and I). The simulation is also implemented in a time-iterative fashion like FDTD. In fact, this method is very similar to the FDTD although originating from different ideas. An interesting comparison between them was made in [24]. Again, there are some commercial simulation packages based on this method, the most notable is MicroStripes produced by Flomerics [25]. An evaluation version can be downloaded from relevant websites.

6.2.5 Comparison of Numerical Methods

There are many other numerical methods (e.g. boundary element methods [26]) which are also used for electromagnetic simulations but are less popular (this does not mean less accurate) than the ones we have discussed. All these numerical methods need to divide the structure of interest into many cells/elements; some approximations are then used to convert the problem into an easily solvable one (such as a matrix equation). A common problem of these

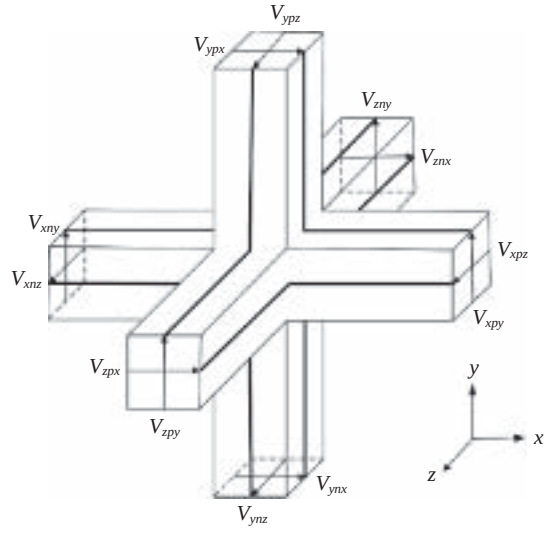


Figure 6.11 A TLM node for simulation

approaches is the huge demand on computer memory and computation time. The accuracy of most commercial packages is quite comparable; the computation efficiency and storage requirements are varied, depending on specific problems. A comparison of frequency and time domain methods is given in Table 6.1. Every method and every software package has its advantages and disadvantages but none is perfect. When these methods are applied to antenna analysis, care has to be taken. For example, the maximum dimension of each cell/element should be smaller than $\lambda/20$ to ensure the accuracy of the results. Some methods (such as the MoM) can easily deal with very thin wire problems whilst other methods find it difficult to handle such problems. When choosing a method or software, all these aspects should be taken into account.

Table 6.1 Comparison of the time and frequency domain methods

	Frequency domain methods		Time domain methods	
	MoM	FEM	FDTD	TLM
Advantages	Fast at single frequency; Easily combined with other methods to deal with large problems.		Broadband results in one simulation; Good for pulse-type problems.	
Disadvantages	Difficult to deal with pulse-type problems;		Not suitable for electrically large systems	
Note	Most suitable for wire-type antennas	Be careful with very thin wires	Be careful with the boundary conditions	Be careful with thin wires

6.2.6 High-Frequency Methods

Numerical methods are generally of good accuracy and flexibility for various configurations. When applied to electrically large structures, they all have problems such as computer memory, runtime or convergence – this is true even for very well-written commercial software based on these methods. The modeling is at present too computationally demanding to be useful for such problems. However, high-frequency methods [27] are particularly suitable for electrically large structures where the wave nature of the radio wave need not be considered. Methods used in optics, a well-developed subject, can be employed for electromagnetic and antenna analysis. These high-frequency methods are usually divided into field-based methods and current-based methods.

Geometrical optics (GO) or *ray optics* is a field-based method and describes light propagation in terms of 'rays' which are perpendicular to the wave fronts of the actual optical waves [27, 28]. GO provides rules for propagating these rays through an optical system, which indicates how the actual wave front will propagate. The basic idea is that the field at distance R away from the source can be obtained using the field in the same tube of rays at distance L away from the source, i.e.

$$|E(R)| = |E(L)| \frac{L}{R} \tag{6.37}$$

This agrees with the conclusion we have obtained for the antenna far field where the field is inversely proportional to the distance. A lot of problems may then be solved by a ray-tracing method.

It should be pointed out that GO is a significant simplification of optics and fails to account for many important optical effects such as diffraction and polarization. Sometimes it is inadequate to completely describe the behavior of the EM field and it is necessary to include the diffracted field – this theory is known as the *geometrical theory of diffraction (GTD)* and the theory for wedge diffraction is called the *uniform theory of diffraction (UTD)*. The combination of GO/GTD/UTD has been applied to large antenna analysis and simulation for many years [29, 30]. Hybrid methods using GO/GTD/UTD with numerical methods (e.g. the MoM) have successfully developed to deal with complex large problems such as antennas mounted on the surface of an electrically large structure [31, 32, 33].

Physical optics (PO) or *wave optics* is a current-based method, building on Huygens's principle (which was discussed in earlier chapters), and models the propagation of complex wave fronts through propagation channels, including both the amplitude and the phase of the wave. This technique can account for diffraction, interference and polarization effects, as well as aberrations and other complex effects, thus it is more general than GO. Equivalent current sources in the illuminated regions are obtained using the equivalence principle introduced in Chapter 3. For a perfect conductor, the equivalent surface current is

$$\boldsymbol{J}_e = \begin{cases} \hat{\boldsymbol{n}} \times \boldsymbol{H}; & \text{in the illuminated region} \\ 0; & \text{in the shadowed region} \end{cases} \tag{6.38}$$

Once the current is known, the radiated field can be found using an equation such as Equation (4.3). Approximations are still generally used when applied numerically on a computer.

Just like GO, PO has also inherited limitations on dealing with diffraction. The *physical theory of diffraction (PTD)* has been developed as an extension of PO that refines the PO surface field approximation just as GTD or UTD refine the GO surface field approximation; more information can be found in [27], for example.

Typical applications of high-frequency methods include the calculation of scattering and diffraction, radar cross-section (RCS), the effects of a finite antenna ground plane, electrically large antennas (such as reflector antennas) and the interaction between antennas and mounting structures.

It is apparent that high-frequency methods are very useful for the analysis of electrically large structures, which may not be the antennas themselves but the mounting structures. These methods offer the following advantages:

- there is no limitation on the maximum dimension of the structure for the simulation, nor is there any runtime overhead in increasing the frequency for the same dimensions of the structure;
- they are not memory or runtime intensive;
- they can compute the interaction of an antenna with a structure which takes over from numerical methods when these become unusable.

6.3 Examples of Computer-Aided Design and Analysis

A large number of electromagnetic modeling software packages have been developed and are available on the market. Some are more successful technically and commercially than others. Some have become industrial standard design and analysis software. It is not possible to give an exhaustive list of the software available (it is a fast-moving industry). In this book we can only deal with some of the popular ones. Since these packages have been developed by many people for many years, they are now well written and normally have a very user-friendly interface for the designer to input the design, and the simulated results can be well presented in 2D or even 3D graphs. There is no point in an individual researcher/engineer developing another general simulation package and reinventing the wheel. However, there is always demand for specialized software, which may perform better than the general purpose software for specific applications.

In this section we are going to use two commercial software packages with different complexity and features to design various antennas. It is hoped that, through these design examples, the reader will gain a better understanding of antenna theory and be able to use some industrial standard software to aid antenna design. Most software suppliers nowadays offer free evaluation packages, which have limited functionality or a limited period of time for the license, but they are good learning tools, enabling one to experiment with designing some simple antennas. The reader is encouraged to explore various packages before any purchase is made. This book does not compare commercial packages and does not offer any shopping advice.

6.3.1 Wire-type Antenna Design and Analysis

As mentioned earlier, the MoM-based NEC program was one of the earliest computer programs developed for wire-type antenna modeling and design. Similar Windows-based commercial

packages are readily available. EZNEC [34] and MININEC [35] are just two well-known examples. Demonstration/evaluation versions, which are basic with a very limited number of segments and restricted functionality, are available for free download. The full versions offer excellent capability (over 1000 segments) at an affordable price (EZNEC cost US$89 and MININEC Professional cost US$790 in 2007). They are great tools for beginners, amateurs and even professionals.

In this subsection we are going to use the EZNEC demonstration version as our design tool, to introduce the major features of this kind of software and to see how it can be used to aid the design and analysis of wire-type antennas.

6.3.1.1 EZNEC

EZNEC is a MoM-based powerful but very easy-to-use program for modeling and analyzing various (especially wire-type) antennas in their intended operating environment. It was developed by Roy Lewallen (W7EL) and has evolved over the years [34]. This Windows-based program is very user friendly and there is a pull-down menu to provide help whenever needed. There are also some examples and a test drive exercise provided with the software package. It takes just a few hours for a new user to become familiar with the program. The major features of this software are:

- The antenna is constructed by straight-wire conductors, thus a surface structure is approximated with wire meshes.
- The input of the antenna is done via a spreadsheet-like entry, the start and end coordinates are required, so is the number of segments.
- The simulation set-up is very straightforward and many short-cuts are built in.
- The computation is very fast and efficient. The results are accurate enough for most wire-type antennas.
- 2D current distribution and 2D and 3D plots of radiation patterns are available; other information such as the gain, input impedance, VSWR, 3 dB beamwidth, front-to-back ratio, take-off angle and side lobe characteristics are also readily available.
- If required, transmission lines, a realistic ground and loads to simulate loading coils, traps or similar components can be selected and added to the simulation.
- Antenna descriptions and pattern plots are easily saved and recalled for future analysis. Multiple patterns can be superimposed on a single graph for comparison. You can see the pattern and antenna currents on the same colour 3D display as the antenna, you can also rotate the antenna display and zoom in for details.

6.3.1.2 Design Examples

Example 6.2: 14 MHz dipole for Ham radio transceiver. 14 MHz (20 m band) is one of the amateur radio (ham radio) bands widely used around the world. Assume that you are going to make a dipole as a transceiver antenna which will be placed in your garden. Use EZNEC to design the antenna and analyze the effects of the ground plane on the antenna input impedance, gain and radiation pattern. Horizontal polarization is assumed.

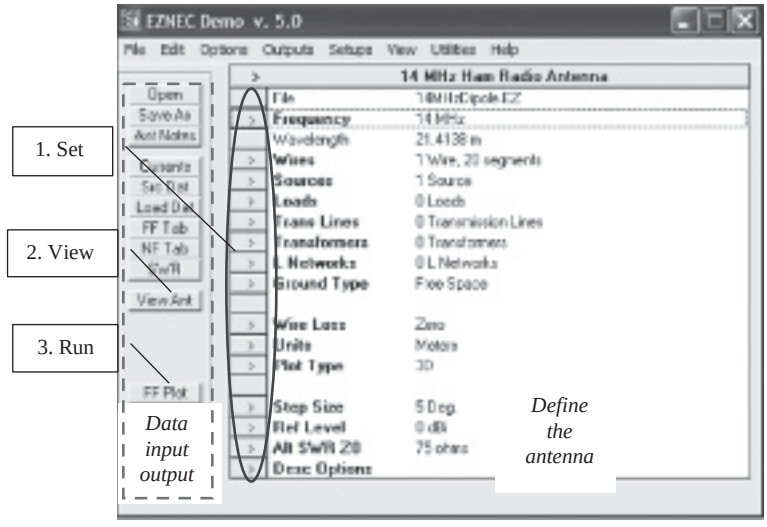


Figure 6.12 EZNEC user interface (Src: source; FF: far field; NF: near field)

Solution:

14 MHz has a wavelength of 21.4 m. To make a resonant and efficient dipole, the dipole length L should be close to $\lambda/2$, more precisely, $L \approx 0.48\lambda \approx 10$ m, which is affected by the ground.

Now let's use EZNEC Demo V.50 (it is free!) to aid the analysis and design. On opening the software, we see the user interface, as shown in Figure 6.12, which is basically divided into two areas: the one on the left is for data input and output and the other one (in the center) is to define the antenna to be simulated. The simulation should follow these steps: define the antenna $\Rightarrow$ view the input antenna $\Rightarrow$ run the simulation $\Rightarrow$ check the results.

The main options to set up the simulation are:

1. *Frequency*: this means the central one, 14 MHz in our case.
2. *Units*: there are five options: m, mm, ft, in and wavelength; we choose 'm'.
3. *Wires* that form the antenna: this means entering the end coordinates of each straight wire, the diameter and the number of segments. You can even input the coating material if applicable, as seen in Figure 6.13. Each segment should be smaller than $\lambda/20$. A wire of 10 m is divided into 20 segments as the first attempt for our design. Clicking on '*View*

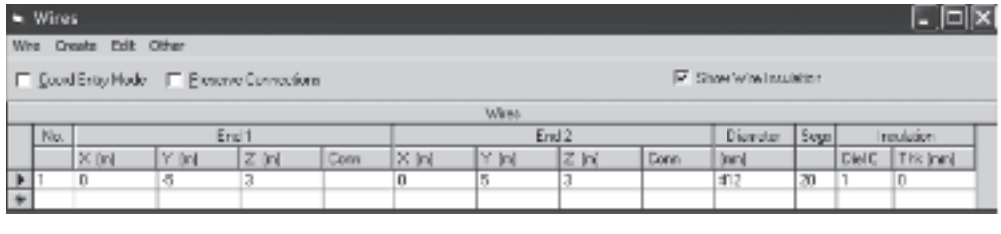


Figure 6.13 Wires input interface

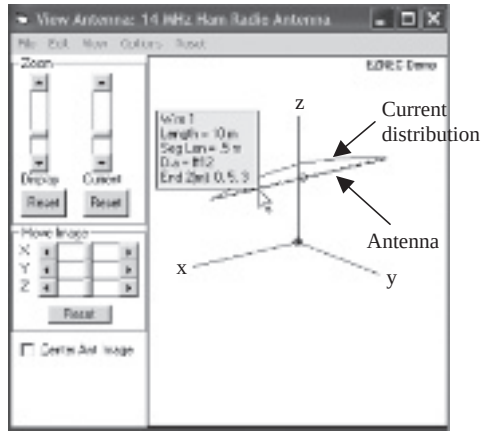


Figure 6.14 Antenna view showing controls and current distribution

antenna', gives Figure 6.14, which shows the antenna (3 m above the ground floor $z = 0$) and the coordinates.

4. *Sources*: the source location, amplitude and phase. Our selection is shown in Figure 6.15. For a center-fed case, it would be better to choose an odd number of segments (such as 19), so the actual position is the specified one.
5. *Loads*: the load impedance at a specific location if applicable.
6. *Transmission line*: location, length, characteristic impedance and loss if used.
7. *Transformer and L Networks* for matching: the port locations and impedances.
8. *Ground type*: there are three options: free space, perfect conducting ground and real ground, which may be defined by the user.
9. *Wire loss*: you can choose a predefined material (such as copper or a perfect conductor with zero loss) or a user-defined one.
10. *Plot type*: 2D (azimuth or elevation) or 3D.

To analyze the effects of the ground plane, we are going to choose three different grounds: free space, a perfect conductor and real ground.

a) **Free space**. First, assuming that the antenna is in free space, we need to select the ground type as '*free space*'. Click on '*SWR*' to define the sweep frequency and then click on '*FF plot*' to start the simulation.

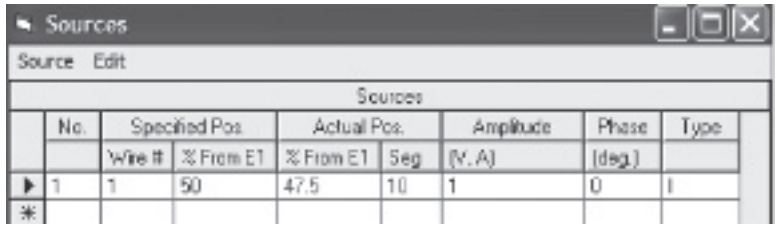


Figure 6.15 Source input interface

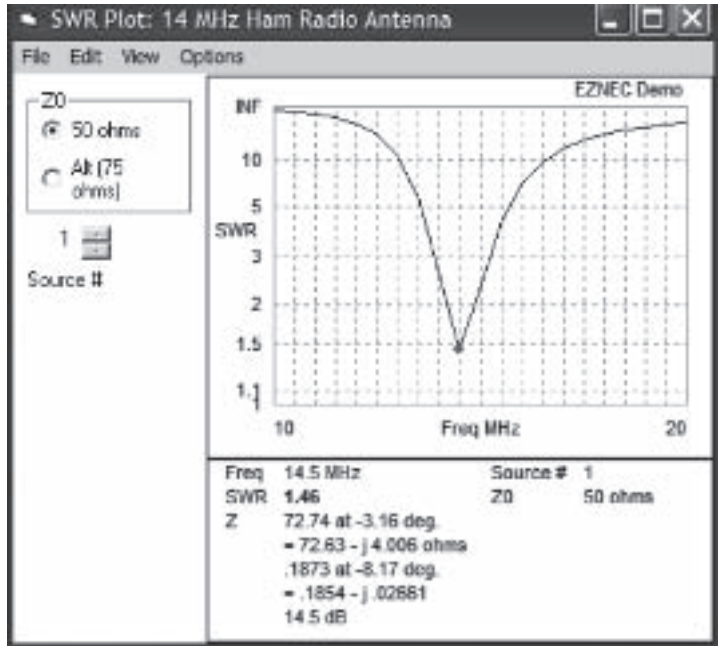


Figure 6.16 VSWR for a 10 m dipole in free space

The results for 10 to 20 MHz are given in Figure 6.16. It is shown that this 10 m dipole resonates around 14.5 MHz with an input impedance of $72.63 - j4.006$ ohms. When it is connected to a 50 ohm transmission line, the VSWR is about 1.46. These numerical results (as well as the radiation pattern in Figure 6.18(a)) are in excellent agreement with our theoretical ones given in Chapter 5.

b) **Perfect ground.** Now we assume that the antenna is 3 m above a perfect ground plane and run the simulation again to obtain Figure 6.17. This shows that the resonant frequency is now shifted down to 14.25 MHz, and the input impedance at 14.5 MHz is changed to $42.87 + j31.02$ ohms. The comparison of radiation patterns for both cases is presented in Figure 6.18. The gain is increased from 2.11 dBi for free space to 8.58 dBi for this case. Thus, the effects of the ground plane are significant – it acts as a mirror to create another antenna (image), forming a two-element dipole array. This array has increased the gain by a factor of (not 2 but) about 4 (6.4 dB)! This is because the radiation pattern is changed from omnidirectional to unidirectional in the H-plane (elevation). It may not look much like the dipole patterns you may have seen (the textbook pictures usually show a dipole in free space), but it is the kind of pattern seen in reality.

c) **Real ground.** If we choose the 'real ground plane', which has a conductivity of 0.005 S/m and relative permittivity of 13, the simulation results show that the resonant frequency is around 14.35 MHz and the input impedance at 14.5 MHz is $61.74 + j14.65$ ohms. The radiation pattern is similar to Figure 6.18(b) and the gain is down to 5.38 dBi while the HPBW is 110 degrees, which is the largest among these three cases. Thus, we can slightly increase the length to 10.2 m (by trial and error) to make it resonate at 14 MHz with a $VSWR = 1.25$.

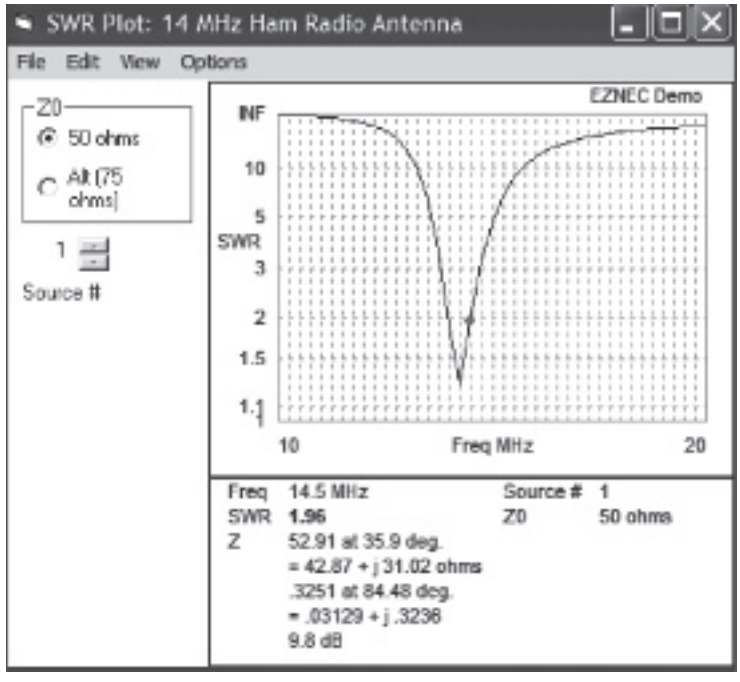


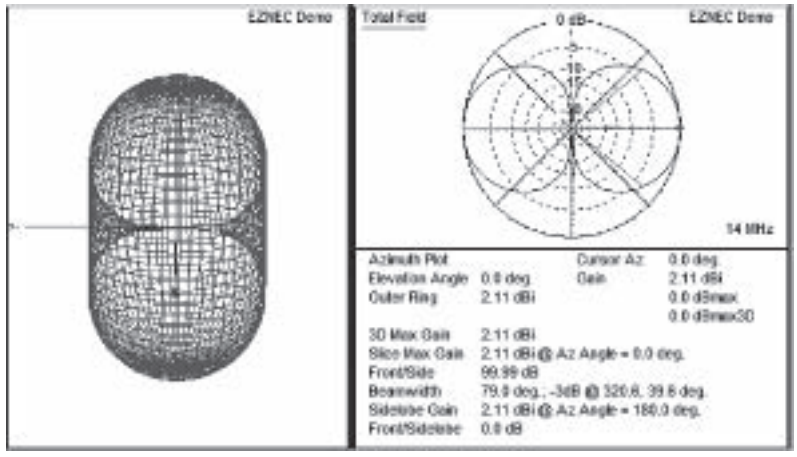
Figure 6.17 VSWR for a 10 m dipole 3 m above a perfect ground

The maximum direction of radiation is at $\theta = 0$, but in practice the wave does not arrive at that angle but more likely at around 45degrees, therefore we can change the height of the antenna to tune the radiation pattern. This tuning process can be guided by the antenna array theory introduced in Chapter 5, since the ground acts as a mirror and this becomes a two-element antenna array. As shown in Figure 6.19, when the antenna is above the ground plane by about half of the wavelength ($z = 9$ m in the figure), the 3 dB radiation pattern covers the elevation angle roughly from 15 to 60 degrees, which matches well with our requirement. The maximum gain is 6.15 dBi, which is larger than that of a quarter-wave monopole (5.15 dBi) – here it can be considered an array.

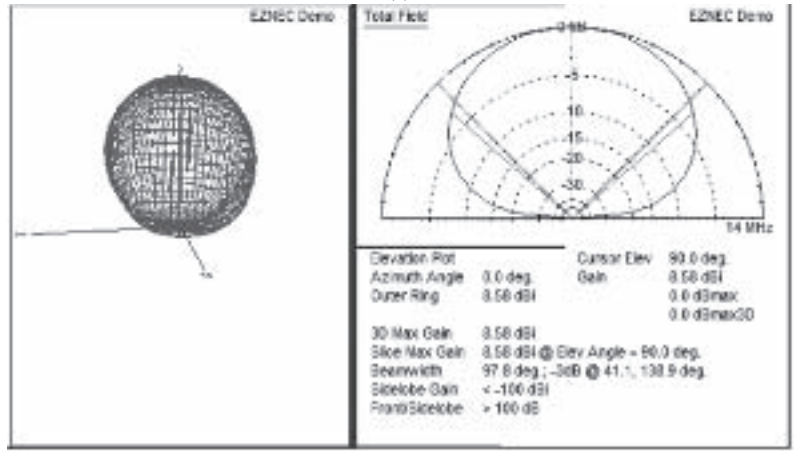
Example 6.3: Monopole array. Two quarter-wave monopoles are separated by a quarter-wavelength. They are placed above a perfect ground plane and fed with the same amplitude and variable phase. Analyze how the phase difference in the feed affects the array performance and compare with theoretical results.

Solution:

For convenience, we choose the frequency to be 300 MHz (so the wavelength is 1 m) and assume the array is made of perfect conducting wire with a diameter of 2 mm. Each quarter-wave monopole is divided into eight segments ($< \lambda/20$). There are now two sources placedat the


EZNEC Demo
Total Field
EZNEC Demo
14 MHz
Azimuth Plot
Elevation Angle 0.0 deg
Outer Ring 2.11 dBi
Cursor Az 0.0 deg
Gain 2.11 dBi
3D Max Gain 2.11 dBi
Slice Max Gain 2.11 dBi @ Az Angle = 0.0 deg.
Front/Side 99.99 dB
Beamwidth 79.8 deg.; -3dB @ 320.6, 39.6 deg.
Sidelobe Gain 2.11 dBi @ Az Angle = 180.0 deg.
Front/Sidelobe 0.0 dB

(a)


EZNEC Demo
Total Field
EZNEC Demo
14 MHz
Elevation Plot
Azimuth Angle 0.0 deg
Outer Ring 8.58 dBi
Cursor Elev 90.0 deg
Gain 8.58 dBi
3D Max Gain 8.58 dBi
Slice Max Gain 8.58 dBi @ Elev Angle = 90.0 deg.
Beamwidth 97.8 deg.; -3dB @ 41.1, 138.9 deg.
Sidelobe Gain < -100 dBi
Front/Sidelobe > 100 dB

(b)

Figure 6.18 Comparison of radiation patterns of a dipole with different grounds (a) 3D and azimuth plot for a free space case (omnidirectional in elevation); (b) 3D and elevation plot for a perfect ground case

ends of the monopoles. A perfect ground is chosen as suggested. The EZNEC input interface and the antenna plot with current distribution are shown in Figure 6.20.

Run the program for the initial phase difference $\varphi_0 = 0,\ \pi/2,\ \pi,$ and $3\pi/2$ respectively to obtain the radiation patterns and the input impedances (click on '*SWR*'). 3D and elevation patterns at $\phi = 0$ are shown in Figure 6.21 and the input impedances are given in Table 6.2. It took just few minutes on a laptop computer to yield all these results, but it could take a much longer time to obtain the theoretical results, especially when the mutual coupling is taken into account.

It is apparent that the feed phase difference between the two monopoles affects both the radiation pattern and input impedance significantly. When the phase difference is 0° (the

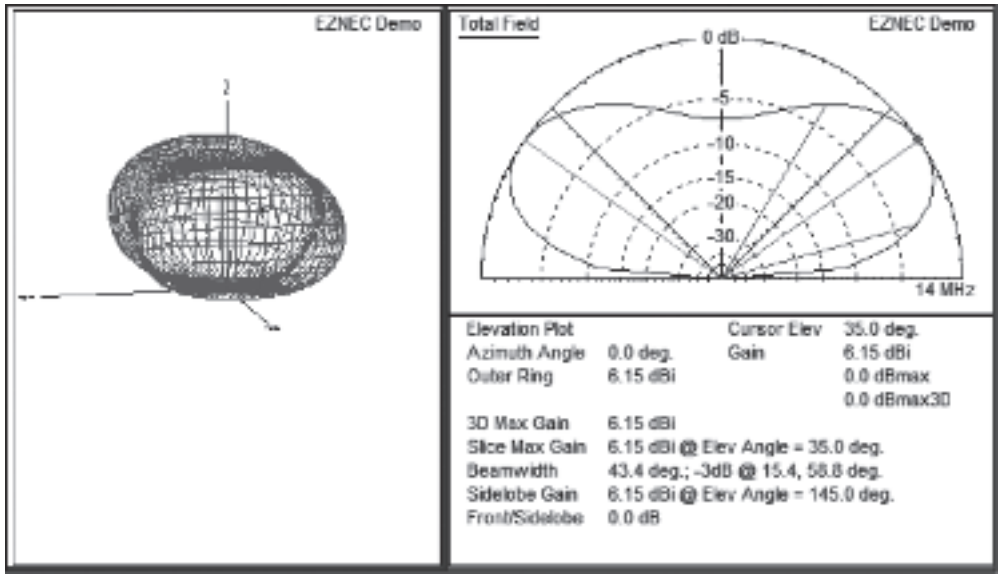


Figure 6.19 Radiation pattern for a dipole of 10.2 m placed 9 m above a real ground plane

same as 360°) or 180°, the radiation pattern is symmetrical about the ZY plane and the two impedances of the monopoles are the same but not 37 ohms, the self-impedance of a λ/4 monopole. In other cases, the pattern is not symmetrical and the two impedances are not the same – these results might be a surprise to some people but they are real in practice. This is further proof of how useful and important the simulation tool can be!

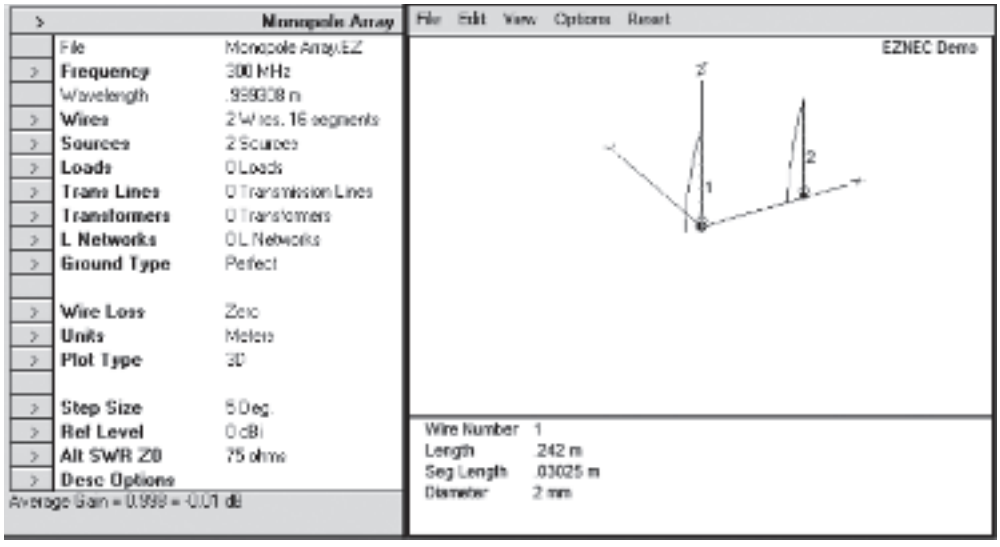
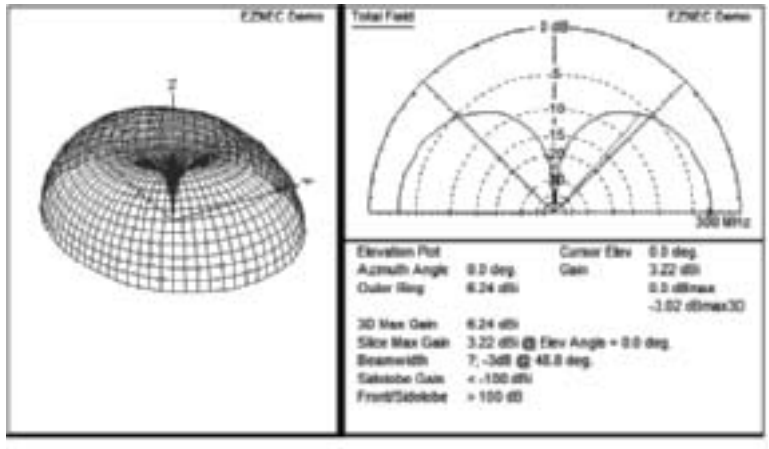
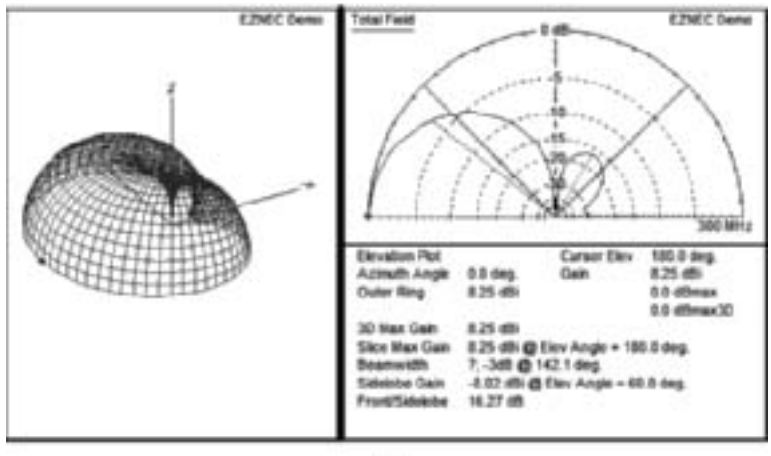


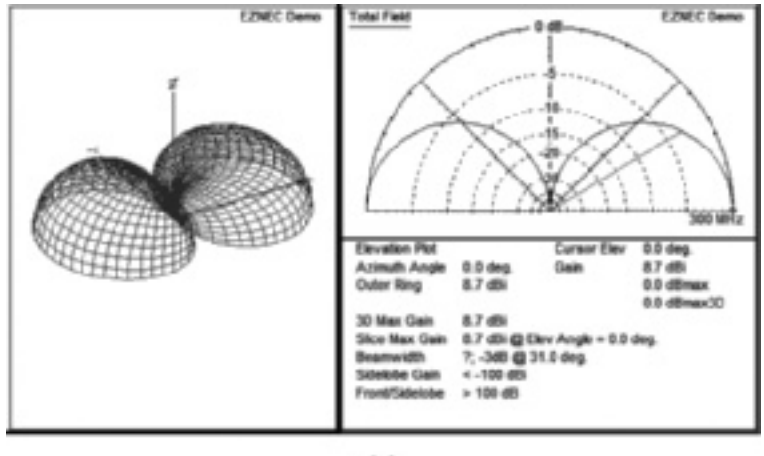
Figure 6.20 Two-monopole array and the EZNEC input



(a)

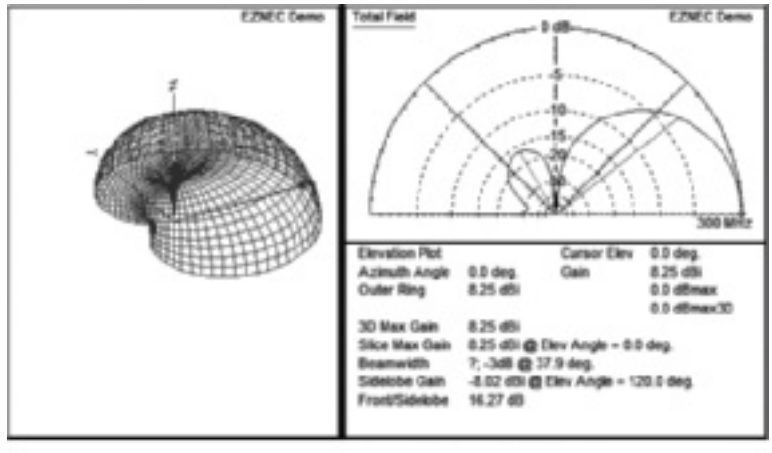


(b)



(c)

Figure 6.21 The radiation patterns of the array with different phase differences (a) $\varphi_0 = 0$; (b) $\varphi_0 = 90^\circ$; (c) $\varphi_0 = 180^\circ$; (d) $\varphi_0 = 270^\circ$



(d)

Figure 6.21 (*Continued*)

It is normally hard to find an analytical solution to the input impedance of an antenna array, but it could be relatively easy to obtain its radiation pattern. We are now going to derive the radiation pattern using the array theory learnt in Section 5.3.

From Equation (5.104), the array factor for this two-element array can be expressed as

$$AF_n = \frac{1}{2}\left[\frac{\sin(2\Psi/2)}{\sin(\Psi/2)}\right] = \cos(\Psi/2) \tag{6.39}$$

where $\Psi = \beta d \sin(\pi/2 - \phi) + \varphi_0 = \pi \sin(\pi/2 - \phi)/2 + \varphi_0$ at the azimuth plane, while $\Psi = \pi \sin(\theta)/2 + \varphi_0$ at the elevation plane. Since the quarter-wave monopole antenna pattern is known (using Equation (5.4) and letting $0 < \theta < 90^\circ$ and also shown in Table 5.2), we can employ the principle of pattern multiplication to obtain the monopole array pattern:

$$E_\theta \propto \left(\frac{\cos(0.5\pi \cos\theta)}{\sin\theta}\right) AF_n \tag{6.40}$$

The antenna array patterns at the azimuth and elevation planes for the initial phase difference $\varphi_0 = 0^\circ,\ 90^\circ,\ 180^\circ,$ and 270° are plotted in Figure 6.22, which are very close to the numerical

Table 6.2 Input impedance for the two monopoles at 300 MHz

Phase difference φ_0	Monopole 1 (ohms)	Monopole 2 (ohms)
0°	$62.74 + j6.188$	$62.74 + j6.188$
90°	$59.4 + j46.21$	$22.76 + j2.846$
180°	$19.38 + j42.87$	$19.38 + j42.87$
270°	$22.76 + j2.846$	$59.4 + j46.21$
360°	$62.74 + j6.188$	$62.74 + j6.188$

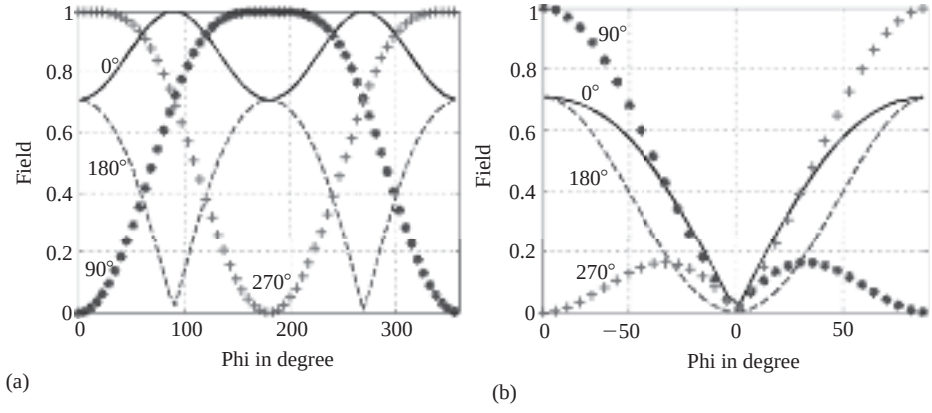


Figure 6.22 Array patterns in the two principal planes for different φ_0 (a) azimuth plane; (b) elevation plane

results presented in Figure 6.21. The only noticeable difference is that, at $\varphi_0 = 90^\circ$, the electric field is zero in theory but very small (-30 dBi) from EZNEC for $\phi = 0^\circ$ and $\theta = 90^\circ$; and at $\varphi_0 = 270^\circ$, we see the same difference for $\phi = 180^\circ$ and $\theta = -90^\circ$. These differences can be explained since, in the theoretical model, the currents on the monopoles are assumed to be the same (as that of a standalone one), but they are not in reality (due to coupling), when they are fed with phase differences (thus, the impedances are different, as shown in Table 6.2). The numerical model has taken the difference into account and may therefore provide more accurate results than the theoretical one with approximations/simplifications.

This type of software is not just limited to wire-type antennas and can also be used for much more complicated simulations. For example, a helicopter simulation model is shown in Figure 6.23 [34], where the surface is replaced by wire meshes and each wire should be smaller than $\lambda/20$. The computation could become a problem if the frequency of interest is too high. In such a case, an alternative approach/software is required.

6.3.2 General Antenna Design and Analysis

Some relatively simple software packages (such as EZNEC) are cheap and very useful for wire-type antenna analysis and design but they are not very suitable for dealing with antennas which have complicated configurations and use various construction materials. More sophisticated

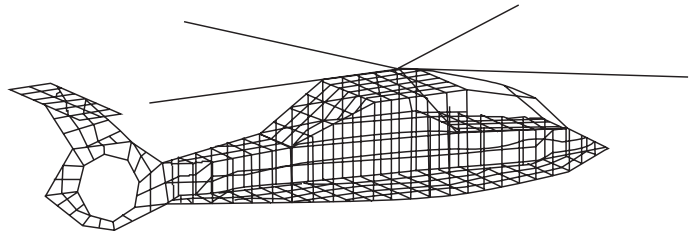


Figure 6.23 A wire model of a helicopter

software packages (such as IE3D, CST Microwave Studio, MicroStripes, FEKO, XFDTD and HFSS, more details can be found from their websites) are very powerful but not cheap. In this section we are going to use Ansoft's HFSS [36] as an example to design a dual-band patch antenna for (GSM) mobile radio communications.

6.3.2.1 High-Frequency Structure Simulator (HFSS)

HFSS utilizes a 3D full-wave finite element method (FEM) to compute the circuit and field behaviors of high-frequency and high-speed components. With HFSS, engineers can extract circuit parameters (such as the impedance), visualize 3D electromagnetic fields (near and far fields) and generate full-wave SPICE^{TM} models to evaluate signal quality effectively, including transmission pathlosses, reflection loss due to impedance mismatches, parasitic coupling and radiation. HFSS can be dynamically linked to other Ansoft software to create a powerful electromagnetic-based design flow and is one of the industrial standard antenna design software packages. The major features include:

- suitability for almost all structures and configurations (wire or nonwire-types);
- all antenna results are given, and some animated results may also be shown;
- ease of optimizing the design (using its parametric function or Optimetrics software);
- good accuracy.

The software is user friendly and relatively easy to operate, but you need to have adequate knowledge of setting up the simulation and interpreting the results. Most of all, you need to have your design ready before simulation – the software does not do the design for you but only helps you to validate and improve your design!

Figure 6.24 shows the user interface, which is basically divided into five main regions: the project manager, the history tree, the property window, the 3D modeler window and the progress window.

6.3.2.2 Design Examples

Now we are going to use the following example to illustrate how to use this software to aid our antenna design.

Example 6.4: Dual-band GSM antenna. Planar inverted F antennas (PIFAs) have been widely reported and are popular in mobile phones. The main reasons for this are: they are of low profile; their radiation patterns are near omnidirectional; they are installed above the phone circuitry, 'reusing' the space within the phone to some degree; they exhibit a low specific absorption rate (and less loss to the head). The objective of this exercise is to design a dual-band PIFA for a GSM mobile handset. It should cover 880–960 MHz (E-GSM 900) and 1710–1880 MHz (DCS 1800 or GSM 1800). The dimensions should be small enough for a standard mobile handset.

Solution:
PIFAs have evolved from the inverted F antenna, as shown in Figure 5.7. For an inverted F antenna, the total length of the antenna should be close to a quarter-wavelength. In our case, the two central frequencies are: 920 MHz ($\lambda_1 = 326$ mm) and 1755 MHz ($\lambda_2 = 171$ mm), we

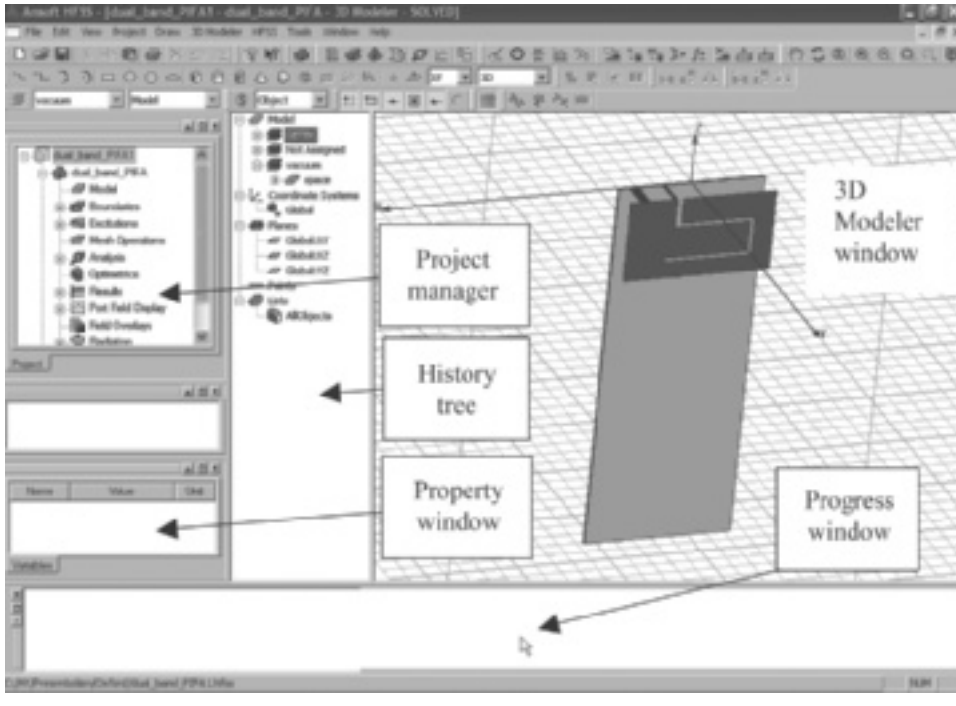


Figure 6.24 The user interface of HFSS

therefore need to create two resonant paths on the antenna: one is about $l_1 = 81.5$ mm and the other is approximately $l_2 = 42.7$ mm. After careful consideration, we decide to use a PIFA antenna of dimensions 40 mm $\times$ 22 mm with a slot to create two current paths which are close to l_1 and l_2 respectively, as shown in Figure 6.25. The detailed initial design is given in Figure 6.26. The PCB size is 40 mm $\times$ 100 mm.

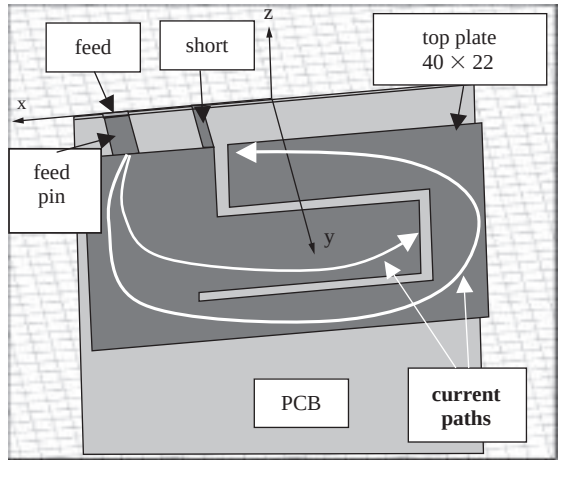


Figure 6.25 A dual-band PIFA antenna

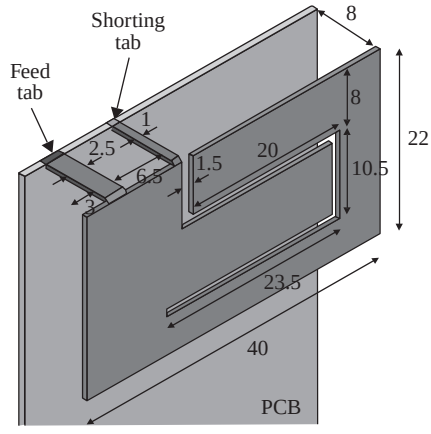


Figure 6.26 A dual-band PIFA (dimensions in mm, slot width constant at 1 mm, apart from where indicated to be 1.5 mm, thickness of antenna plate 0.5 mm)

It is almost impossible to use EZNEC to simulate such a structure. However, it is relatively straightforward to employ software such as HFSS to model it. The step-by-step procedures of using HFSS to simulate the antenna are:

1. Launch HFSS: you should see a schematic user interface as in Figure 6.24.
2. Set tool options and open a new project if required.
3. Create the 3D antenna model: this is the main part of entering your designed antenna. You need to input the coordinates of every corner. The detailed dimensions are given in Figure 6.26.
4. Assign boundaries to all surfaces: there are many options, such as radiation, perfect E, and perfect H.
5. Assign excitations: there are also many types, 50 ohm lumped port is selected for this antenna analysis.
6. Set up an analysis: you need to select the solution frequency, frequency sweep if required and the accuracy requirement and number of passes. For our simulation the frequency is set from 800 MHz to 2000 MHz to cover the desired frequency range.
7. Validate the model: to check if there are any errors in the inputted model.
8. Start the simulation: the real-time progress report is shown in the progress window.
9. View solution data: you should view the convergence first to see if the results have converged, then create reports on various antenna parameters. If the results have not converged, more iterations are required. The requirement on the computational power (memory and time) could be excessive. This is a major and common problem for such a simulation.

For our design, the convergence plot is shown in Figure 6.27 and the results have converged after eight iterations. The simulation used 250 MB memory and 320 MB disk space. The return loss is illustrated by Figure 6.28. As expected, there are two resonant bands: one is around 910 MHz and the other one is near 1750 MHz. Figure 6.29 shows the radiation pattern at 1755 MHz, which is close to omnidirectional and has a gain/directivity of 4.358 dBi.

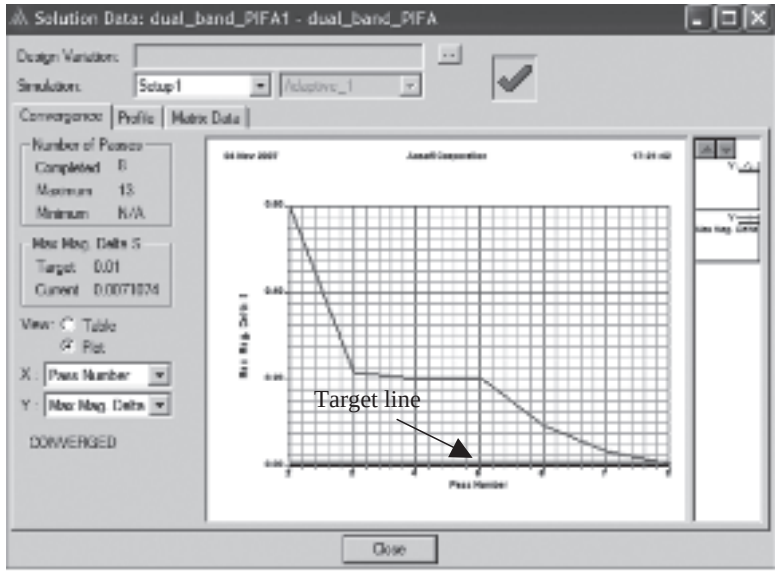


Figure 6.27 The convergence plot

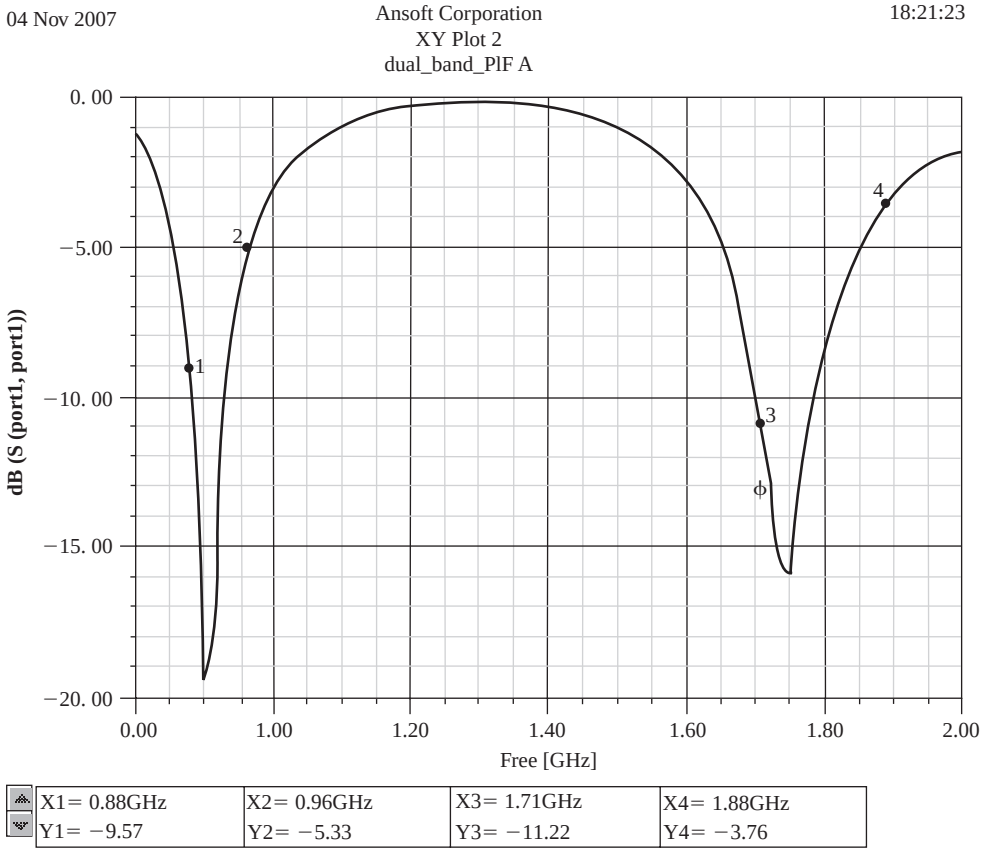


Figure 6.28 S_{11} in dB as a function of the frequency in GHz

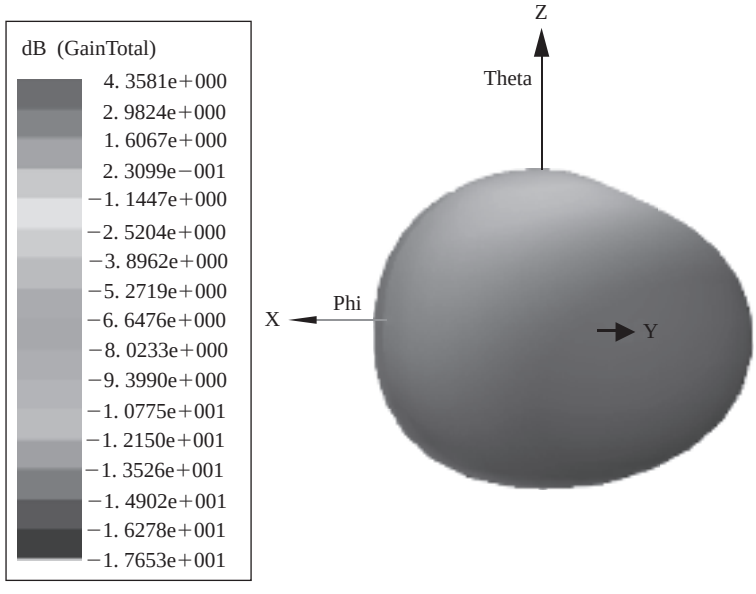


Figure 6.29 Radiation pattern at 1755 MHz

To further assist our analysis, we can plot other important information, such as the current distribution at 1755 MHz on the plate, as shown in Figure 6.30 along with the finite element meshes which were generated adaptively by the software. We can also obtain the animated current/field on the antenna. It is seen that the current is concentrated along the edge of the slot and the pattern is a standing wave at 1755 MHz (resonant). This corresponds to the shorter path identified in Figure 6.25.

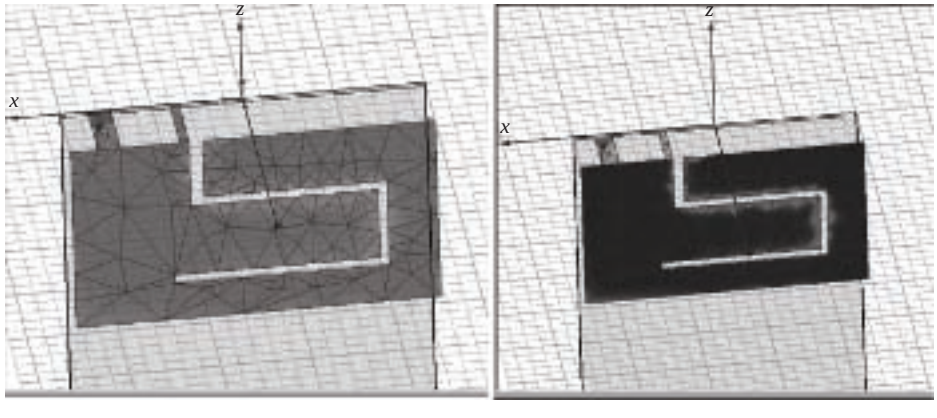


Figure 6.30 Meshes and current distribution on the antenna

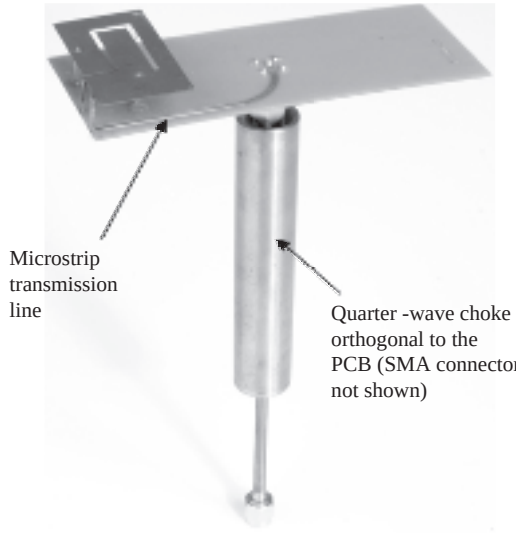


Figure 6.31 The dual-band PIFA antenna with feed line and RF choke

For comparison, an antenna with the designed dimensions has been made, as shown in Figure 6.31. Due to the special driving point of this antenna, a microstrip line of 50 ohms is created as its feed line, which is connected to an SMA connector. A quarter-wave choke is employed to minimize the effects of the measurement cable. The transmission line that runs from the central connector to the antenna driving point is de-embedded from the measured results. More detailed information about this antenna can be found in [37] and general discussions on how to make accurate antenna measurements are presented in the next chapter.

The simulated and measured impedances of this configuration over the 800–1040 MHz and 1610–2090 MHz bands are shown in Figure 6.32(a) and (b), respectively. Two measured results are shown, which are based on the same configuration and measured using the same techniques to give an indication of the manufacturing tolerances involved.

The simulated and measured reflection coefficients (S_{11}) in dB are presented in Figure 6.33. For the resonant frequency, simulations and measurements tie up very well in the 800 MHz band. In the 1800 MHz band, differences in the resonant frequency of approximately 15 MHz are observed. At both frequency bands, the resonant frequency simulation accuracy is thought to be of the same order as the manufacturing accuracy. Some phase rotation is apparent at both the 800 MHz band and, to a lesser degree, the 1800 MHz band. The 1800 MHz measurements have a noticeably higher resistance than the simulations, although this is not observed at GSM. It is expected that manufacturing accuracy can be improved in future work. Residual differences between simulations and measurements are then expected to depend on de-embedding inaccuracies and simulation errors.

There are many other useful functions which may be employed to aid our design. For example, the design for Example 6.4 does not completely meet the industrial return-loss requirement ($>$ 6 dB) over the 1800 MHz band. Fine tuning is required: this can be achieved

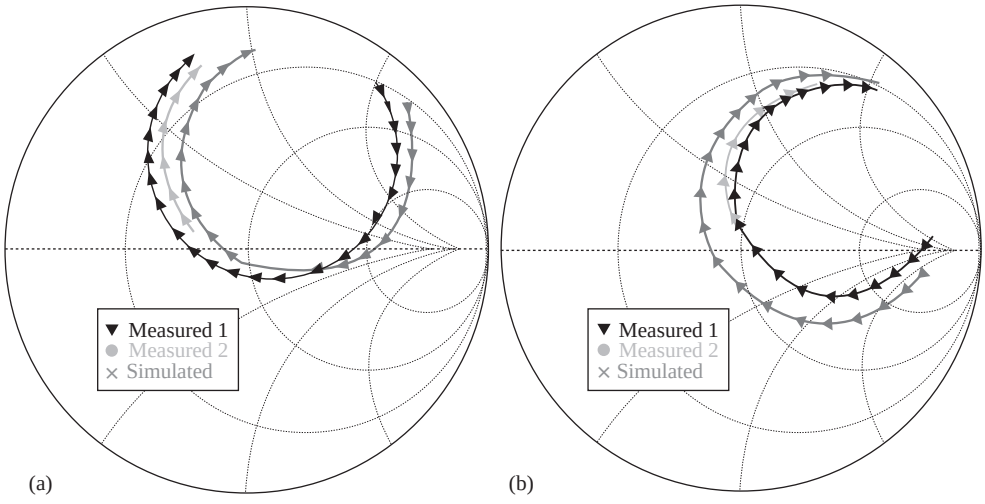


Figure 6.32 Simulated and measured impedances on a Smith Chart (a) impedance over 800–1040 MHz; (b) impedance over 1610–2090 MHz

by using the optimization function of the software. We can set a parameter (such as the width of the antenna) as the variable to be optimized and run the computer software to find the optimum value for this parameter. Details on this kind of advanced feature can be found from the software suppliers and are not covered by this book.

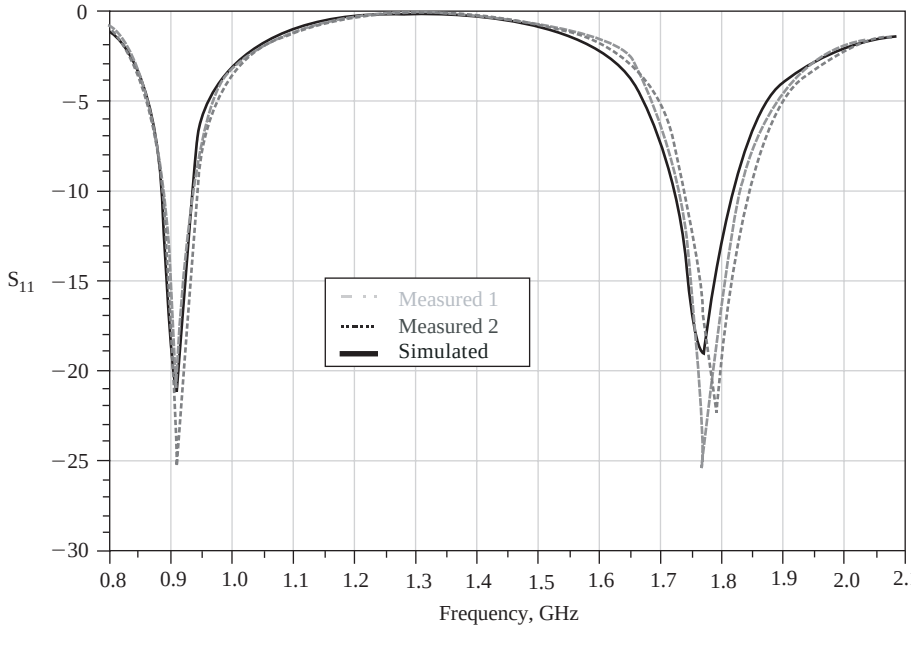


Figure 6.33 Simulated and measured S_{11} in dB

6.4 Summary

In this chapter we have introduced the computational electromagnetic techniques used in modern antenna design. The focus has been placed on the MoM, which gave us insight into how a complex field equation could be solved numerically and how to obtain the desired antenna parameters. Two software packages have been employed as examples to illustrate how they can be utilized for antenna design and analysis.

References

[1] G. J. Burke and A. J. Poggio, 'Numerical electromagnetics code (NEC) – method of moments,' *Technical Document 11*, Naval Ocean Systems Center, San Diego, USA, January 1981.

[2] J. Julian, J. M. Logan and J. W. Rockway, 'MININEC: a mini-numerical electromagnetics code,' *Technical Document 516*, Naval Ocean Systems Center, San Diego, USA, 1982.

[3] J. Rockway, J. Logan, D. Tan and S. Li, *The MININEC System: Microcomputer Analysis of Wire Antennas*, Artech House, 1988.

[4] B. Friedman, *Principles and Techniques for Applied Mathematics*, John Wiley & Sons, Inc., New York, 1956.

[5] R. F. Harrington, 'Matrix methods for field problems,' *Proc. IEEE*, **55**(2), 136–149, 1967.

[6] R. F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.

[7] Matlab: http://*www.matlab.com*

[8] C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd edition, John Wiley & Sons, Inc., 1997.

[9] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd edition, John Wiley & Sons, Inc., 1998.

[10] J. H. Richmond, 'Digital computer solutions for rigorous equations for scattering problems,' *Proc. IEEE*, **53**, 796–804, 1965.

[11] G. A. Thiele, 'Wire antennas' in *Computer Techniques for Electromagnetics* by R. Mittra (Ed.), Pergamon Press Ltd, New York, 1973.

[12] A. Hrennikoff, 'Solution of problems of elasticity by the frame-work method,' *ASME J. Appl. Mech*., **8,** A619–A715, 1941.

[13] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, 1973.

[14] J. L. Volakis, A. Chatterjee and L. C. Kempel, *Finite Element Method Electromagnetics: Antennas, Microwave Circuits, and Scattering Applications*, John Wiley & Sons, Inc. and IEEE Press, 1998, p. 368.

[15] O. C. Zienkiewicz and R. L. Taylor, *Finite Element Method, 5th edition, Volume 1 – The Basis*, Elsevier, 2000.

[16] K. Yee, 'Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media', *IEEE Trans. Anten. and Propag*., **14**, 302–307, 1966.

[17] G. Mur, 'Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic field equations,' *IEEE Trans. Electromag. Compat*., **23**, 377–382, 1981.

[18] A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, 1995.

[19] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method, 3rd edition*, Artech House, 2005.

[20] P. B. Johns and R. L. Beule, 'Numerical solution of 2-dimensional scattering problems using a transmission-line method,' *Proc. IEE*, **118**(9), 1203–1208, 1971.

[21] W. J. R. Hoefer, 'The transmission line matrix (TLM) method,' in *Numerical Techniques for Microwave and Millimeter Wave Passive Structures*, by T. Itoh, Ed., 1990.

[22] P. B. Johns, 'A symmetrical condensed node for the TLM-method,' *IEEE Trans. Microwave Theory Tech.*, MTT35(4), 307–311, 1987.

[23] C. Christopoulos, *The Transmission-Line Modelling Method in Electromagnetics*, Morgan and Claypool Publishers, 2006.

[24] M. Krumpholz, C. Huber and P. Russer, **'**A field theoretical comparison of FDTD and TLM,' *IEEE Trans. Microwave Theory Tech.*, **43**(8), 1935–1950, 1995.

[25] MicroStripes: *http://www.flomerics.com/products/microstripes/index.php*

[26] P. K. Banerjee, *The Boundary Element Methods in Engineering*, McGraw-Hill College, 1994.

[27] P. K. Pathak, 'Techniques for high-frequency problems,' in *Antenna Handbook* by Y. T. Lo and S. W. Lee (Eds), Van Nostrand Reinhold Company, New York, 1988.
[28] E. Hecht, *Optics*, 4th edition,. Pearson Education, 2001.
[29] A. R. Lopez, 'The geometrical theory of diffraction applied to antenna and impedance calculation,' *IEEE Trans. Antennas Propag.*, **14**, 40–45, 1966.
[30] G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves*, 3rd edition, Peter Peregrinus Ltd (IEE now IET), 1986.
[31] TICRA: GRASP, *http://www.ticra.com/*
[32] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd edition, John Wiley & Sons, Inc., 1998.
[33] G. A. Thiele, 'Hybrid methods in antenna analysis,' *IEEE Proc*., **80**(1), 66–78, 1992.
[34] EZNEC: *http://www.eznec.com/*
[35] MININEC: *http://www.emsci.com/*
[36] HFSS: *http://www.ansoft.com/*
[37] K. Boyle, M. Udink, A. de Graauw and L. P. Ligthart, 'A Dual-Fed, Self-Diplexing PIFA and RF Front-End,' *IEEE Trans. on AP*, **55**, 373–382, 2007.

Problems

Q6.1 There are many antenna design software packages available free of charge (for evaluation versions). Use the information provided in this chapter as a guide to download at least one software package and get familiar with the software, which will be used for the following exercises.

Q6.2 Using your newly downloaded software, redo Examples 6.2–6.4 and identify if there are any differences in the results.

Q6.3 Use the new software to design a loop antenna which operates at 2.45 GHz and has an omnidirectional radiation pattern.
   a. If the loop is electrically small, find the radiation pattern and input impedance, then compare with the theoretical results.
   b. If the loop is about one wavelength and the desired VSWR is smaller than 2 for a 50 ohm feed line, obtain a good design and justify whether a matching circuit is required.
   c. If the loop is made of a conducting wire, discuss the effects of the diameter of the wire on the loop input impedance and gain.

Q6.4 With the aid of software, design a circularly polarized helix antenna of an end-fire radiation pattern with a directivity of 13 dBi and find out its input impedance, *HPBW*, *AR* and radiation pattern. Then compare these results with those of Example 5.2.

Q6.5 Design a patch antenna specified by Example 5.7 with the aid of computer software and then compare your results with Example 5.7.

Q6.6 The initial design of a six-element Yagi–Uda antenna is given in Table 5.4. With the aid of antenna software, find its input impedance, radiation pattern and gain.

Q6.7 With the aid of software, design a circularly polarized antenna for GPS application. The frequency should cover 1559 to 1610 MHz (L1) band, $VSWR < 2$ for a 50 ohm feed line, and the radiation pattern should be unidirectional (towards the sky)

7

Antenna Manufacturing and Measurements

Once an antenna is designed, it should be made and tested. The construction of an antenna may be a complex process since the antenna has to meet the electrical and mechanical specifications as well as some other requirements (such as costs). In this chapter we are going to see which materials are normally employed to make antennas, what antenna measurements should be conducted and how.

7.1 Antenna Manufacturing

From a construction point of view, antennas (such as dipoles, loops and horns) are normally manufactured using conducting materials, low-loss dielectric materials (e.g. dielectric resonant antennas) or a combination of both (e.g. patch antennas). The selection of the right material and a robust construction are important elements in making a successful antenna.

7.1.1 Conducting Materials

Conducting materials are the most widely used to build antennas. In principle, all conductive materials can be employed, but the antenna efficiency is closely linked to the conductivity of the material: the higher the conductivity, the higher the efficiency. Thus, in practice, we only choose materials with very good conductivity.

In addition to the conductivity, we have to take other things into account, which include the following aspects:

- **Mechanical considerations**: the material should be strong enough to keep its desired shape under normal working conditions.
- **Environmental considerations**: the material should be resilient to environmental changes – for example, it should not oxidize or erode – in order to maintain its good conductivity. Sometimes, a layer of paint is applied to meet this requirement. The paint may affect the antenna performance, so a trade-off is necessary.

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

- **Cost considerations**: as for any other product, the cost of antenna manufacture should be kept to a minimum.
- **Weight consideration**: it should normally be as light as possible.

As a result, copper, brass (an alloy of copper and zinc), bronze (an alloy of copper and tin) and aluminum are widely used to make antennas. Composite materials are also employed for constructing antennas. For example, polycarbonate/acrylonitrile butadiene styrene (PC/ABS) alloys are widely used for making mobile phone antennas. They combine the excellent heat resistance of polycarbonate with the improved processibility of ABS. The conductivities of some common metals used in antenna construction are given in Table 1.3 in Chapter 1.

To construct prototype antennas (especially those of special shapes) at home or in a lab, metallic tapes or even aluminum foils (from a supermarket) may be employed, since they are conductive, very flexible and cheap. However, it is necessary to ensure that the metal thickness is much greater than the skin depth of the material at the desired frequency. A discussion on skin depth was provided in Section 3.6.1.

One of the most overlooked antenna construction considerations is the *galvanic corrosion* that usually occurs when two dissimilar metals are brought into physical contact (mated) during the assembly process and exposed to the weather. There can be serious corrosion at their respective contact points. There is a direct relationship between various types of dissimilar metals when they are mated. Some dissimilar metals (such as copper and brass) when mated cause very little corrosion. There are other metals, however, that react most harshly when mated. Zinc and brass, for example, will cause corrosion, with the zinc metal quickly breaking down. There are several ways of reducing and preventing this form of corrosion. One of them is to electrically insulate the two metals from each other. This can be done using plastic or another insulator to separate the metals. Use of absorbent washers that may retain fluid is often counter-productive. Sometimes they have to be in electrical contact, the best solution is to use the same metal throughout the construction. If this is not possible, the next course of action is to assemble materials that have a close relationship on a galvanic metals table, as shown in Table 7.1.

Table 7.1 Galvanic metals table

Most Active Metal	
	Zinc
	Aluminium
	Iron steel
	Stainless steel
	Lead
	Tin
	Nickel
	Brass
	Copper
	Bronze
	Silver
	Gold
Least Active Metal	

7.1.2 Dielectric Materials

Dielectric materials are employed to form the desired antenna shape (a conductor may be inside the dielectric or the dielectric may be covered by a conductor/composite material), to protect a metal antenna or to act as a dielectric antenna. The dielectric properties of the materials that are commonly used in antenna design are given in Table 1.2 and Table 2.4 (PCB substrates).

Dielectric resonator antennas (DRAs) are a relatively new type of antenna that have some attractive properties and have become more widely used due to the availability of low-loss and high-permittivity materials [1–3]. The typical relative permittivity is from 10 to 100. Because of the discontinuity between the dielectric and air, DRAs act as resonators whose resonant mode is excited by a specially designed feeding method. The coupling of power to the structure can be achieved using coaxial probes, microstrip or coplanar transmission lines. Dielectric resonators of any shape may be used for antenna design. The shape and dielectric properties of the dielectric along with the frequency and excitation determine which resonant mode may be generated, hence the radiation pattern and input impedance. The widely used geometries are rectangles, cylinders, rings and hemispheres. DRAs offer a number of good features, including small size and reasonable bandwidth. The use of low-loss materials and small conductor losses permits high radiation efficiencies. Unlike conductive antennas, the influence of nearby objects (such as human hands) has a limited effect on the performance of a DRA, which is one of the major reasons for DRAs becoming popular in some portable devices.

It should be pointed out that there is normally a trade-off between the antenna size and its bandwidth. The smaller the size (which means a larger permittivity), the narrower the bandwidth. This result is applicable not only to DRAs but also other antennas, such as patch antennas – a high permittivity substrate is required if the antenna has to be made small.

From the antenna design point of view, the main requirement on materials for coating and protection purposes is low loss (in addition to other requirements), which affects the antenna efficiency and input impedance. Another noticeable effect of the dielectric material on the antenna is that the resonant frequency is shifted downwards, which will be demonstrated in Section 7.3.

7.1.3 New Materials for Antennas

Antenna evolution has been heavily influenced by the materials industry. PCBs are commonly used for making planar antennas. There is a trend to integrate antennas with an RF front-end and circuits. Thus, the development of PCBs has an impact on antenna manufacturing. Liquid crystal polymer (LCP) material provides an alternative to traditional polyimide film for use as a substrate in flexible circuit construction. This development offers improvements in manufacturing and processing of LCP flexible circuits and antennas. LCP film is directly metalized using a vacuum sputtering process. Unlike laminated substrates, the sputtered LCP substrate yields a good metal-to-LCP adhesion, avoids trace line undercut problems, eliminates remaining copper issues and provides effective circuit-patterning resolution. Either subtractive or additive processing can be employed, which is good for antenna applications. It has good thermoplastic properties with a low dielectric constant and loss factor remaining constant over the frequency range of 1 kHz to 45 GHz, with a negligible moisture effect. The properties of multilayer (three-dimensional) vertical integration capability, good electrical and mechanical properties and near-hermetic nature make this substrate a practical choice for the design of low-cost antennas [4].

More recently, some new materials and processes have been developed. Silver inks and deposited coppers are employed for making small and lightweight antennas for applications such as RFID [5]. Electrically conductive adhesives are replacing conventional soldering. Artificial materials have been developed and provide some unique features. For example, the reflection coefficient on a high impedance surface structure can be 1 (not -1 as on a good conductor). Some new antennas of low profile and high gain using these new materials are emerging [6, 7, 8]. Although it is quite unlikely that they will become popular for all applications, they do provide opportunities for people to use these materials to design new antennas for some special applications.

7.2 Antenna Measurement Basics

Once an antenna is designed and constructed, it is essential to validate the design with proper measurements, a crucial element of the development process. *The most important measurements are the impedance and radiation pattern measurements*, since they are the most important characteristics of an antenna. The input impedance may be specified as a value at a particular frequency and/or as a maximum VSWR or return loss over a range of frequencies (often referred to 50 ohms). The measurement is relatively straightforward. However, the radiation measurement is much more complicated and time-consuming. Typical measurements are the radiation pattern and gain measurements, which should be performed in an antenna test range. The radiation patterns of some antennas (such as those for radar, microwave links, some cellular base stations and satellites) are highly directional and often have tightly specified envelopes. Not only will the peak gain be specified, but challenging requirements might also be placed on parameters such as polarization purity (axial ratio), side lobe levels and efficiency. The testing and evaluation of some of these parameters may be difficult. Details of impedance measurements are given in Section 7.3. Details of radiation measurements are given in Section 7.4. Some other measurement issues are discussed in Section 7.5.

7.2.1 Scattering Parameters

In previous chapters we defined and used the reflection and transmission coefficients for both the circuit and field problem analyses. These concepts can be generalized for network analysis.

A two-port network problem can be illustrated by Figure 7.1, where a_1 and a_2 are the input whilst b_1 and b_2 are the output at Port 1 and Port 2, respectively. This network is characterized by *scattering parameters*, or *S-parameters*:

$$[\boldsymbol{S}] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{7.1}$$

Figure 7.1 A two-port network

which links the input to the output by

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{7.2}$$

Thus, we have:

$$\begin{aligned} S_{11} &= \text{Port 1 reflection coefficient} = b_1/a_1; \\ S_{12} &= \text{Port 2 to Port 1 transmission coefficient/gain} = b_1/a_2; \\ S_{21} &= \text{Port 1 to Port 2 transmission coefficient/gain} = b_2/a_1; \\ S_{22} &= \text{Port 2 reflection coefficient} = b_2/a_2. \end{aligned} \tag{7.3}$$

S-parameters are actually reflection and transmission coefficients for a network of N ports. In this case, $N = 2$. These parameters were originally introduced in optics, where optical waves were scattered by objects. The concepts were later extended to radio waves and RF engineering, but the term 'S-parameters' has remained unchanged.

It should be pointed out that if a network is passive and contains only isotropic and loss-free materials that influence the transmitted signal, the network will obey

- the *reciprocity principle*, which means $S_{21} = S_{12}$ or more generally $S_{mn} = S_{nm}$, and
- the *law of power conservation*

$$[S]^H [S] = [I] \tag{7.4}$$

where $[S]^H$ is the *complex conjugate transpose* of the S-parameter matrix $[S]$ and $[I]$ is the identity/unit matrix. For a 2-port loss-free network, we have

$$\begin{aligned} |S_{11}|^2 + |S_{21}|^2 &= 1 \\ |S_{22}|^2 + |S_{12}|^2 &= 1 \end{aligned} \tag{7.5}$$

which is equivalent to

$$|a_1|^2 + |a_2|^2 = |b_1|^2 + |b_2|^2 \tag{7.6}$$

The input power is the same as the output power, as expected.

Almost all antennas, attenuators, cables, splitters and combiners are reciprocal (but maybe not loss-free) networks. The networks which include anisotropic materials in the transmission medium, such as those containing ferrite components, will be nonreciprocal. Although it doesn't necessarily contain ferrites, an amplifier is also an example of a nonreciprocal network.

Now the question is how are S-parameters linked to antennas? From Figure 7.2, we can clearly see that a transmitting-receiving antenna system in the space can be considered a 2-port network. The transmission and reflection can be characterized using S-parameters. S_{11} and S_{22} are the reflection coefficients of Antenna 1 and Antenna 2, respectively. They indicate how well the antenna feed line is matched with the antenna. S_{21} and S_{12} are the transmission coefficients from one antenna to another. They are determined by the characteristics of both antennas (such

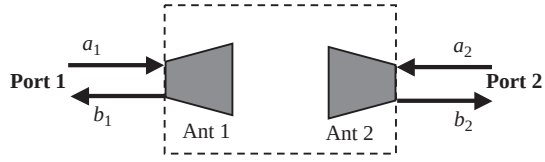


Figure 7.2 The equivalent 2-port network of a transmitting-receiving antenna system

as radiation patterns and matching) and the separation between them. It should be pointed out that this equivalent network is not a confined 2-port network but a lossy open network, since the power radiated from one antenna is not all received by another antenna, thus the conditions in Equations (7.5) and (7.6) do not hold for such an antenna network. However, the reciprocity principle can still be applied to this case.

7.2.2 Network Analyzers

For RF/microwave engineering, the *oscilloscope* is often employed to view signals in the time domain while the *spectrum analyzer* is used to examine signals in the frequency domain. For antenna measurements, the most useful and important piece of equipment is the *network analyzer* (NA), which is basically a combination of a transmitter and a receiver. Normally it has two ports and the signal can be generated or received from either port. The main parameters it measures are the S-parameters, i.e. it measures the reflection and transmission characteristics of a network. There are two types of network analyzer:

- *the scalar network analyzer* (SNA) measures the amplitude of the parameters of a network, such as VSWR, return loss, gain and insertion loss;
- *the vector network analyzer* (VNA) measures both the amplitude and phase of the parameters of a network.

The VNA is much more powerful than the SNA; in addition to the parameters which can be measured by the SNA, it can also measure some very important parameters such as the complex impedance, which is essential for antenna measurements. Thus, we need a VNA not an SNA for antenna measurements.

7.2.2.1 The Configuration of a VNA

A typical VNA is shown in Figure 7.3. It can be seen that the equipment comes with two standard coaxial connectors. The measured results can be saved to a disk or a computer via a USB port at the back. The architecture of a VNA is given in Figure 7.4. This clearly shows that the source signal can be transmitted from either Port 1 or Port 2 to the device under test (DUT), which is controlled by a switch. A part of the signal from the source is provided directly to the reference R, which will be compared with the received signal A (from Port 1) or B (from Port 2) by the central processing unit (CPU).

VNAs can perform most of the necessary measurements without the need for any manual tuning. Just as importantly, they also contain integrated computers and graphical displays,

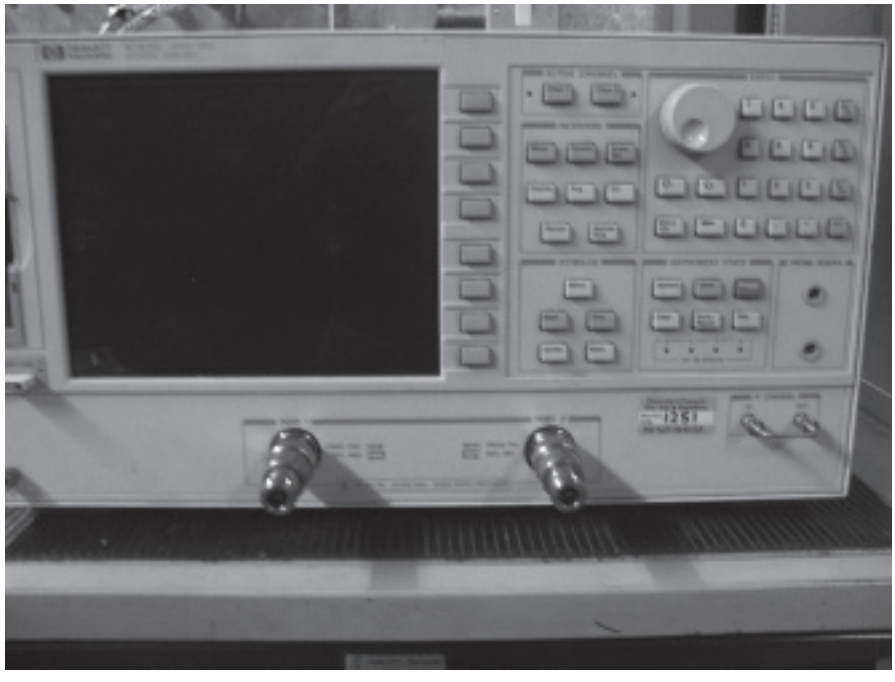


Figure 7.3 A typical VNA

which allow an unprecedented level of data manipulation and display – for example, making it possible to plot calibrated, phase-adjusted impedances directly onto an on-screen Smith Chart. Indeed, network analyzers now have most of the features of typical personal computers: high levels of processing power, high-resolution colour screens, familiar operating systems, network connections, etc. Furthermore, most network analyzers support the standardized file formats that are used by circuit simulators, allowing measurements to be performed and then imported into simulations of the system in which the antennas operate.

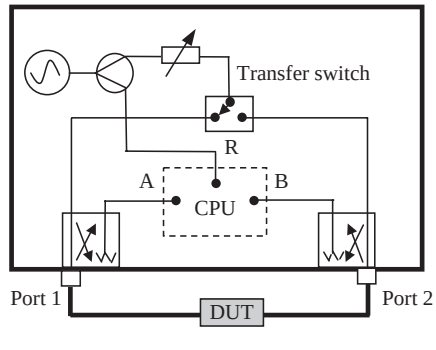


Figure 7.4 Typical configuration of a VNA

7.2.2.2 What Can a VNA be Used to Measure?

The VNA is frequency-domain equipment; it can obtain the signal in the time domain using Fourier transforms. For example, it can be used as a time-domain reflectometer (TDR) to identify discontinuities of an antenna, a transmission line or a circuit. For antenna measurements, the typical parameters that can be measured by a VNA include:

Transmission measurements:
- gain (dB)
- insertion loss (dB)
- insertion phase (degrees)
- transmission coefficients (S_{12}, S_{21})
- electrical length (m)
- electrical delay (s)
- deviation from linear phase (degrees)
- group delay (s)

Reflection measurements:
- return loss (dB)
- reflection coefficients (S_{11}, S_{22})
- reflection coefficients vs distance (Fourier transform)
- impedance ($R + jX$), which can be displayed on the Smith Chart
- VSWR

Note these parameters can be displayed on the VNA screen. However, when the data are exported to a computer or saved to a disk, we normally get the complex S_{11}at each sampling frequency.

It should also be pointed out that, although the VNA can be viewed as the combination of a transmitter and a receiver, the received signal is not displayed as its absolute but a relative value, which is different from a conventional receiver. Thus, the VNA cannot normally be used as a receiver or spectrum analyzer.

7.2.2.3 Calibration and Measurement Errors

When using a VNA to perform antenna measurements, a careful calibration has to be conducted. The reasons are:

- As a radiator, the antenna should not be placed too close to the VNA (to avoid coupling and interference), that is, it should not be directly connected to the VNA. Thus, a cable and connectors have to be used.
- The cable and connectors introduce attenuation and a phase shift.
- The reading on the VNA is at the default reference plane, but what we want to measure is the reading at the input port of the antenna.

We therefore need to remove the effects of the cable and connectors and shift the measurement reference plane right to the end of the cable – this process is called calibration. The standard calibration needs three terminations for one-port calibration, i.e. short, open and load/matched. The commonly used calibration for a 2-port network is short-open-load-thru

(SOLT) calibration. Other calibration methods, such as thru-reflection-line (TRL) calibration [9], short-open-load-reciprocal element (SOLR) and line-reflection-match (LRM, the best calibration accuracy up to 110 GHz) are also used in practice.

There are various errors that may be introduced into the measurement. The well-known ones include system errors, random errors and drift errors. Because of the radiating nature of the antenna, antenna measurements have many other possible sources of error. The system errors can be removed by calibration. Some errors (such as random error) cannot be eliminated. Just like simulation results, measured results are also approximations.

7.3 Impedance, S_{11}, VSWR and Return Loss Measurement

Up until approximately the late 1980s, most impedance measurements were performed with manually tuned impedance bridges. Such bridges are still used today – for example, operating impedance bridges are used for measurements of AM broadcast antennas due to their ability to be connected 'in line' whilst high powers are transmitted by the antenna. However, for laboratory use, the VNA has become the industry standard equipment for high-frequency (distributed element) impedance measurements and it is very different from the multimeter used for DC and low-frequency (lumped element) impedance measurements.

From the circuit theory provided in Chapter 2, we know that the load impedance and the transmission line impedance determine the reflection coefficient, VSWR and return loss. Since the line impedance for the VNA is normally set at 50 ohms, the antenna impedance, S_{11} (defined by Equation (7.3)), and return loss measurements are basically the same when a VNA is employed. We just need to measure S_{11} (a complex number) at one port where the antenna under test (AUT) is connected. The standard measurement procedures are:

- Select a suitable cable (low loss and phase stable) for the measurement and ensure that it is properly connected to the VNA (otherwise a significant error could be generated) – this is a major source of measurement error.
- Select the measurement frequency range and suitable number of measurement points over the frequency.
- Perform the one-port calibration and ensure that the cable is not moved (or errors could be generated).
- Conduct the measurements in an environment with little reflection (such as an open area or an anechoic chamber – to be discussed in Section 7.4.2).
- Record the measured results.

7.3.1 Can I Measure These Parameters in My Office?

In practice, many people perform this measurement in a standard lab or office – this is fine provided that the antenna is not very directional and the signal reflected may be considered small. The most sensitive frequency band is around the resonance. For a very directional antenna, one may point the antenna toward a RAM (radio absorbing material) to reduce the reflection.

Let us take the dual-band PIFA antenna designed in Example 6.4 as an example. A comparison of the simulated and measured results was provided in Chapter 6. Here the measured S_{11} (in dB) in an office (dashed line) and in an anechoic chamber (solid line) are shown in

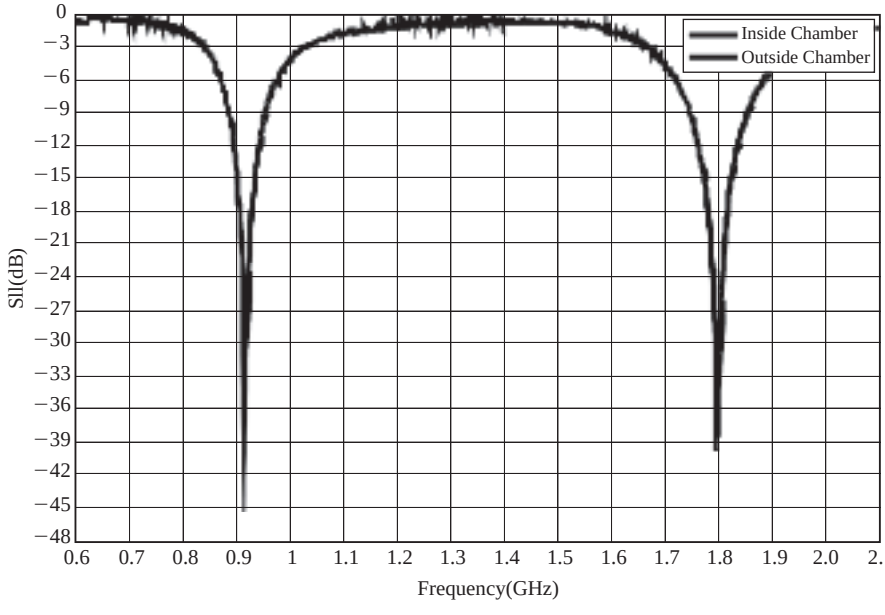


Figure 7.5 Results comparison of a dual-band antenna measured in an office and in an anechoic chamber

Figure 7.5. It is apparent that the difference between them is too small to notice, since the directivity of this antenna is small. The only noticeable difference is indeed around the two resonant frequencies (the two troughs).

7.3.2 Effects of a Small Section of a Transmission Line or a Connector

In order to make measurements, very often we have to use a connector and probably a small section of a transmission line that cannot be included in the calibration. As a result, they may introduce errors to the measured results. Their effects are more significant on the impedance than on the VSWR and return loss. This is because the impedance is a complex number and is sensitive to a length change of the line, while the VSWR and return loss are scalar and not sensitive to a length change.

7.3.3 Effects of Packages on Antennas

The effects of a ground plane and the installation on antennas were discussed earlier and it was concluded that both the impedance and radiation pattern might be affected. Here we use the dual-band PIFA antenna again as an example to illustrate how the package shown in Figure 7.6 may affect the antenna resonance. The package case is made of a plastic material with low loss over the frequency range of interest. The measured results of S_{11} in dB for the antenna with (solid line) and without (dashed line) the package case are shown in Figure 7.7, which is from 0.7 to 1.2 GHz for the GSM 900 band. It is apparent that the resonant frequency is shifted down from 0.965 to about 0.905 GHz, which is quite considerable. Thus, its package and installation must be taken into account when the antenna is designed.

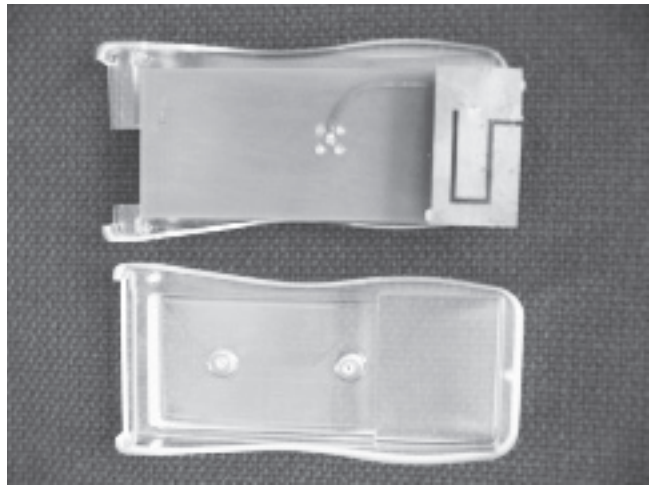


Figure 7.6 A dual-band mobile antenna with its package case

7.4 Radiation Pattern Measurements

As one of the most important parameters, the radiation pattern has to be measured once an antenna is developed, which is a much more difficult and time-consuming task than the impedance measurement. There are a few methods of measuring antenna radiation patterns. They can be classified as near-field and far-field measurement methods, or frequency domain

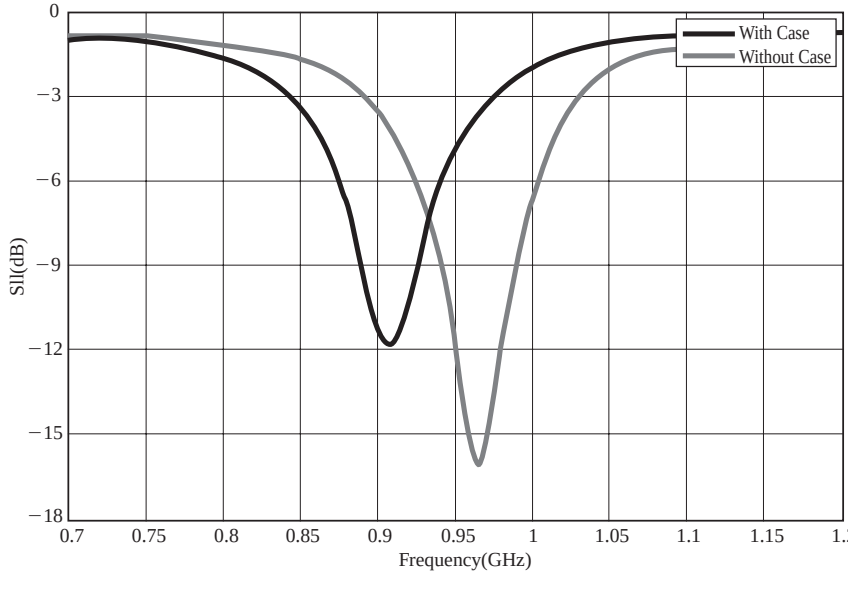


Figure 7.7 The measure S_{11} (in dB) with (solid line) and without (dashed line) the case over 0.7 to 1.2 GHz

and time domain methods [10–14]. Which method is chosen depends primarily on the antenna size and location. For example, medium wave broadcast antennas are installed on steel towers. It is possible to measure in the hemisphere above the ground, but a helicopter or hot-air balloon would be required to do so: normally this would be prohibitively expensive. Instead, measurements are normally performed on the ground along 'radials' that emanate at the antenna phase center and extend outwards (with chosen angle increments) for a few tens of kilometers. Measurements are made at various distances along these radials (at points where access to land is allowed) and radiation patterns are calculated from these measurements, where possible making allowances for the conductivity of the ground. This is something of an extreme example, but it illustrates that the method of antenna measurement can be very specific to a particular antenna type. In this section we will only deal with some common pattern measurement methods, focusing on some of the most widespread applications.

7.4.1 Far-Field Condition

Antenna pattern measurements are usually performed in the far field: defined, somewhat nominally, as the separation required between source and test antenna such that the phase variation of the wave front across the aperture of the test antenna is less than $\pi/8$ radians (22.5 degrees), as illustrated in Figure 7.8. In addition to the AUT, a source antenna is required to generate the far field.

Referring to Figure 7.8, it is clear that the circular wave front from the phase center of the source antenna gives rise to a phase variation across the aperture of the test antenna, of width D, given by

$$\Delta\phi = \left(\frac{2\pi}{\lambda}\left(\sqrt{l^2 + \frac{D^2}{4}} - l\right)\right) = \left(\frac{2\pi l}{\lambda}\left(\sqrt{1 + \frac{D^2}{4l^2}} - 1\right)\right) \tag{7.7}$$

Assuming that $l \gg D/2$, this becomes

$$\Delta\varphi \approx \frac{\pi D^2}{4\lambda l} \tag{7.8}$$

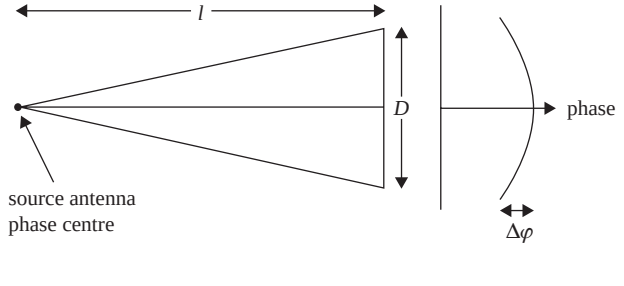


Figure 7.8 Phase variation across an antenna aperture

Defining the far field as the separation, l, required for a phase variation of less than $\pi/8$ across the aperture gives

$$l \geq \frac{2D^2}{\lambda} \tag{7.9}$$

It is very important to note:

- As discussed in Section 3.1.1, Equation (7.9) is for electrically large antennas. For smaller antennas, $l > 3\lambda$ is also required.
- When the antenna is rotated, the distance from one end to the source varies from $l + D/2$ to $l - D/2$. The $1/r$ field dependence due to this rotation can give rise to pattern errors. If a ± 0.5 dB amplitude error is acceptable then $l > 10D$ is required.

7.4.2 Open-Area Test Sites (OATS)

Open-area test sites (OATS) are, as the name suggests, outdoor sites where no reflectors, other than the ground, are present over a relatively wide area. The most attractive advantage of OATS is the low cost. A typical site is illustrated in Figure 7.9.

It can be seen from Figure 7.9 that the path difference between the direct and reflected waves is

$$\Delta = \left(\sqrt{l^2 + (h_T + h_R)^2} - \sqrt{l^2 + (h_T - h_R)^2}\right) \tag{7.10}$$

where

l: distance between the transmit and receive antennas
h_T: height of the transmit antenna above ground
h_R: height of the receive antenna above ground

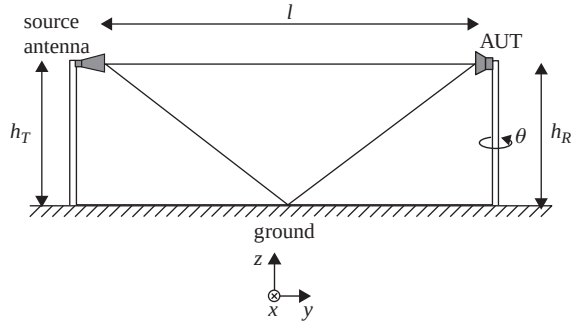


Figure 7.9 Open-area test site (OATS)

Using the binomial expansion and the assumption that $l \gg (h_T + h_R)$ gives

$$\left[1 + \left(\frac{h_T \pm h_R}{l}\right)^2\right]^{0.5} \approx 1 + \frac{1}{2}\left(\frac{h_T \pm h_R}{l}\right)^2 \tag{7.11}$$

Substituting in Equation (7.10) gives

$$\Delta \approx \frac{2\,h_T\,h_R}{l} \tag{7.12}$$

If we assume that $l \gg (h_T + h_R)$, we can neglect the distance-dependent amplitude differences between the two waves and assume that both the direct and reflected components travel over the distance l. We must, however, account for amplitude and phase differences due to the reflection coefficient of the ground. Since the phase term of the wave is $e^{j\beta \cdot r}$, the signal at the receiver is given by

$$E = E_D\left[1 + \Gamma \exp\left\{j\frac{4\pi h_T H_R}{\lambda l}\right\}\right] \tag{7.13}$$

where E_D is the field strength that would be measured at the receiver in free space. Hence, using the Friis transmission formula in Chapter 4, the received power can be written as

$$P_R = P_T\left(\frac{\lambda}{4\pi l}\right)^2 G_T G_R \left|1 + \Gamma \exp\left\{j\frac{4\pi h_T h_R}{\lambda l}\right\}\right|^2 \tag{7.14}$$

Obviously, if the reflection coefficient of the ground is not zero, the received power at a fixed distance l is not uniform. For example, if it is a metal ground plane, the horizontal and vertical reflection coefficients approach –1. This gives

$$P_R = 2P_T\left(\frac{\lambda}{4\pi l}\right)^2 G_T G_R \left\{1 - \cos\left(\frac{4\pi h_T h_R}{\lambda l}\right)\right\} \tag{7.15}$$

which can be simplified to

$$\frac{P_R}{P_T} = 4\left(\frac{\lambda}{4\pi l}\right)^2 G_T G_R \sin^2\left(\frac{2\pi h_T h_R}{\lambda l}\right) \tag{7.16}$$

When $\lambda l \gg 2\pi h_T h_R$, this equation can be further simplified to Equation (3.45) – the two-ray model result. But, generally speaking, the received power is a periodic function of $h_T h_R$ as a periodic function. With the presence of ground reflection, the power at the AUT is sensitive to the heights of transmit and receive antennas, their separation and the wavelength. The field is not suitable for pattern measurements. Thus, it is very important to minimize the reflection from the ground plane by raising the antenna height or placing RF-absorbing materials (RAM – more details to follow in the next section) between the two antennas to allow accurate measurements.

There are other issues to be considered when using OATS. For example, the site must be located in a place where the ambient level of RF interference is low. This is increasingly

difficult due to the density of wireless communications systems and locating a site in a remote area (in terms of radio coverage) causes logistical problems associated with the transport of devices and personnel. Since OATS are outdoors, testing time is limited by daylight hours and weather conditions. In some countries there is very little daylight during cold winters; often such seasonal variations are inconsistent with business requirements. To counter the problems associated with testing time, measures may be taken to extend the number and duration of operational days. However, weather shields, lighting (and the associated cables) and heating systems are potential sources of reflections and interference.

7.4.3 Anechoic Chambers

Anechoic chambers were developed so that antenna engineers could enjoy the convenience and productivity of being able to perform measurements indoors. The main obstacle to this is that 'indoors' implies 'within walls' and the walls reflect antenna signals and, hence, distort radiation pattern measurements. The term *anechoic* simply means without echoes: measures are taken – usually using RAM – to prevent reflections from the walls, ceiling and floor. In addition, the walls are usually metalized to prevent external interference from entering the measurement room. A typical anechoic chamber is illustrated in Figure 7.10.

The chamber consists of a metallic room with a source antenna at one end that is used to excite an AUT at the other end. The antenna under test is usually mounted on a positioner (or turntable) that is rotated in the azimuth plane (the $x - y$ plane in Figure 7.10) to obtain a two-dimensional radiation pattern. Some chambers may be equipped with positioners that are also capable of movement, or *tilt*, in the elevation plane, whilst others may have so-called *roll over azimuth* positioners in which the AUT rotates orthogonally to the azimuth rotation, allowing measurement of the full 3D radiation pattern. The source antenna may be rotated about its axis for polarization measurements. The chamber is usually arranged to transmit from the source antenna and receive on the AUT, although this may be reversed if it is more convenient (the reciprocity principle). The turntable positioner, plotter, receiver and transmit source are all under computer control.

Some chambers are rectangular in shape; some chambers are tapered from the source to the AUT to prevent the formation of standing waves and to make the wave front incident on the AUT more planar, increasing the effective chamber length.

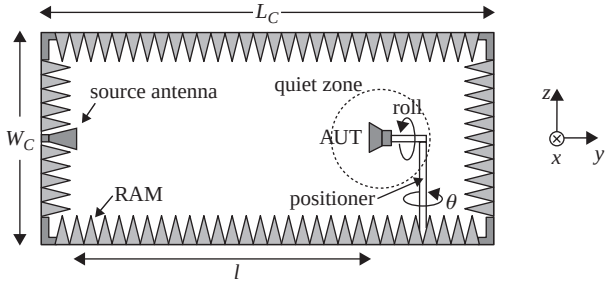


Figure 7.10 Typical anechoic chamber

Radiation pattern measurement accuracy is limited by the finite reflectivity of the chamber walls, the positioner and the cables that are used to feed the AUT. Multiple reflections 'fill-in' nulls in the radiation pattern and create false side lobes. The shape of the room is designed to minimize reflections from the walls over a wide range of incident angles and frequencies. The lowest frequency of operation is determined by the length of the absorber – typically the absorber needs to be approximately one wavelength long at the lower end of the operational range. At 300 MHz – the lower end of the UHF band – this corresponds to one meter, which is too large for installations in normal commercial buildings. The upper end of the frequency range is determined by the composition of the absorbers and their surfaces. Most RAM is made of carbon-loaded polyurethane foam (for the resistivity and reduced weight). Pyramids and wedges are widely used shapes. Pyramids work best at normal incident. A tapered impedance transition from the free space to the back of the absorber ensures broadband-absorbing performance.

The dynamic range of an antenna chamber is limited by the source transmit power and receiver sensitivity. This is not a serious limitation provided the signal is detected in a narrow bandwidth and a wide dynamic range low-noise amplifier is used at the test antenna. Often a network analyzer can be used, perhaps with an additional transmit amplifier.

At low frequencies (< 1 GHz, say) an open test range would normally be used. This is because any chamber would need to be very long to operate in the far field and, also, a large volume of RAM would be needed, since the RAM depth should be a couple of meters or more for low frequencies. However, some very large chambers exist that are capable of operation at lower frequencies and near-field measurements can also be used to reduce the measurable frequency of indoor systems.

Anechoic chambers can be used to measure antenna radiation either in the near or far fields. Near-field measurements are dealt with in the following two sections.

7.4.3.1 An Example of an Anechoic Chamber

A picture of an anechoic chamber at the University of Liverpool is shown in Figure 7.11. The chamber has dimensions of 6.0 m $\times$ 3.5 m $\times$ 3 m and most of the RAM has a length of 300 mm for frequencies from 1 GHz to 40 GHz with a return loss $>$ 40 dB. The source antenna is a broadband double-ridged horn antenna, and the quiet AUT area is about 0.5 m $\times$ 0.5 m $\times$ 0.5 m for an antenna separation of 3 m, which is suitable for the measurement of most mobile radio antennas. The system set-up is given in Figure 7.12; a VNA is used as the transmitter and receiver. A computer-controlled turntable/positioner (which has two step motors for azimuth and elevation controls) is connected to a controller via an optical link (to minimize measurement errors). This automatic measurement system and its software were developed by Diamond Engineering [15] at a very reasonable price, which is particularly suitable for small antenna measurements.

7.4.4 Compact Antenna Test Ranges (CATR)

For some electrically large antennas, the far-field distance given in Equation (7.9) is too large for a conventional anechoic chamber or even an OATS; a possible solution is to utilize a compact antenna test range (CATR) to generate a plane-wave-like uniform AUT area. The

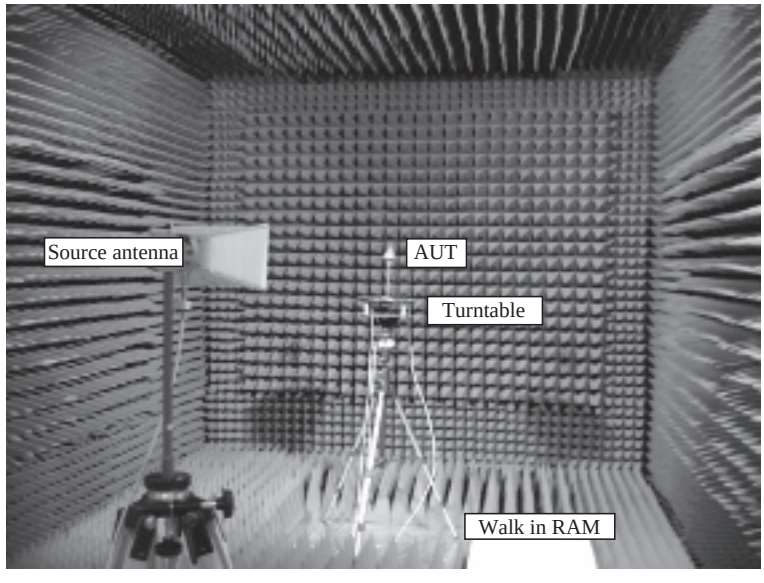


Figure 7.11 An example of an anechoic chamber

idea of CATR is simple: just replace the source antenna by a large offset parabolic reflector antenna (as discussed in Section 5.2.3) in a conventional anechoic chamber or RATS. The AUT should be smaller than the reflector. It has been reported that quiet zones encompassing 50% of the reflector width can be accomplished. More details can be found in [13, 16].

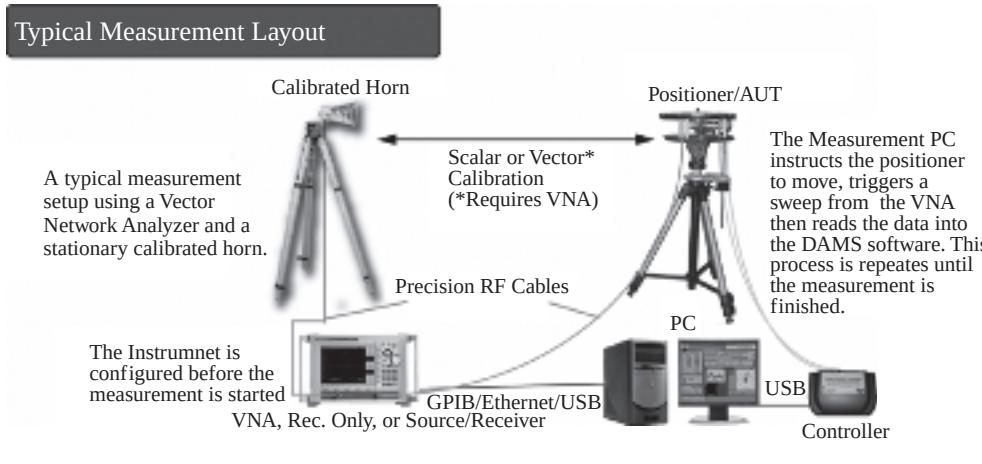

Typical Measurement Layout
Calibrated Horn
Positioner/AUT
A typical measurement setup using a Vector Network Analyzer and a stationary calibrated horn.
Scalar or Vector* Calibration (*Requires VNA)
The Measurement PC instructs the positioner to move, triggers a sweep from the VNA then reads the data into the DAMS software. This process is repeates until the measurement is finished.
Precision RF Cables
PC
The Instrumnet is configured before the measurement is started
GPIB/Ethernet/USB
USB
VNA, Rec. Only, or Source/Receiver
Controller

Figure 7.12 Automatic antenna measurement system set-up. (Reproduced by permission of Diamond Engineering)

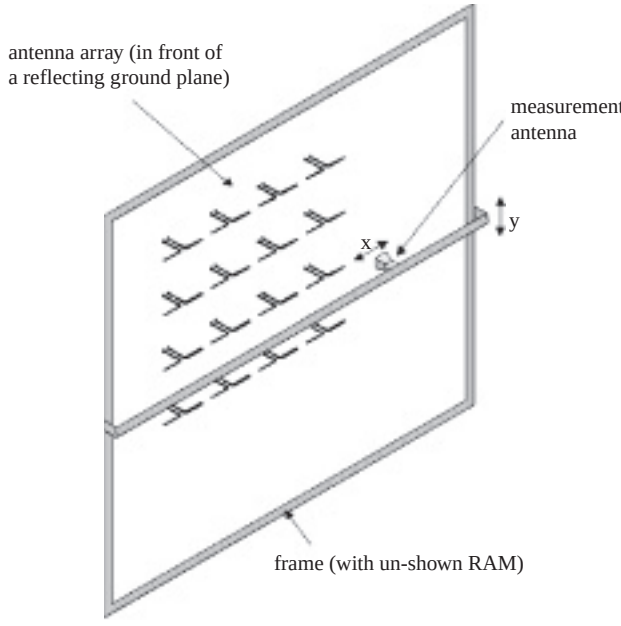


Figure 7.13 Planar near-field chamber

7.4.5 Planar and Cylindrical Near-Field Chambers

Planar and cylindrical near-field chambers are particularly useful for measuring aperture antennas and arrays. The former, shown schematically in Figure 7.13, is best at measuring planar arrays (with many elements in the x and y dimensions), particularly those with a reflector for unidirectional radiation. The latter is best for measuring arrays with more elements in the y than the x dimension and with omnidirectional radiation in the azimuth plane.

The system shown in Figure 7.13 operates by moving a measurement antenna – or probe – across a plane close to and extending beyond the aperture of the antenna. The spacing between this plane and the antenna under test is large enough to prevent mutual interaction with the probe, but small enough to allow a compact system. The probe scans in the antenna near field and a mathematical transformation is applied to give far-field radiation patterns.

7.4.6 Spherical Near-Field Chambers

Spherical near-field chambers have the advantage of measuring the entire field around the antenna: this allows the total power emanating from the antenna to be measured and, hence, the efficiency calculated. This makes such systems particularly suitable for the measurement of handset antennas, where efficiency is especially important (since handset antennas operate in a multipath environment, receiving signals from all directions, the gain in any one of these directions is relatively unimportant, since an averaging effect generally occurs).

Spherical near-field measurement systems generally fall into two categories: swinging-arm systems and fixed-probe systems. Swinging-arm systems are less complicated electrically

than fixed-probe systems, but are slower due to the mechanical movement that is necessary. Probe-based systems have been made possible (at a competitive cost) due to the computational power of modern computers. This allows each probe to be compensated digitally every time a measurement is made (for all angles and frequencies).

A typical commercial system is shown in Figure 7.14. The chamber is large enough that it is possible to do measurements of handset antennas with the complete system present – i.e. the phone and the user. The scanning method is fast enough to allow users to stay still for the duration of the measurement (measurement of the full sphere at an angular resolution of approximately 5 degrees with 20 frequency points is possible in approximately 3 minutes).

The system consists of a circular array of dual-polarized wideband measurement probes (visible as crosses in Figure 7.14), for elevation measurement and a rotating turntable for azimuth measurement. The diameter of the circle of probes is approximately 3.2 m. This, combined with the near-field measurement technique, allows measurement of large radiating structures. The speed of the system is attributed to the method used to switch between probes. The method is termed the *modulated scattering technique*: each probe is modulated 'on' while all others are modulated 'off'. All of the probes are connected via a passive combining network

Figure 7.14 A probe-based spherical near-field measurement SATIMO system

to an RF unit, which then separates the 'on' probe from the 'off' probes in the frequency domain. This process allows the probes to be sequentially scanned very quickly, without the need for discrete switches.

The system shown operates from 450 MHz – 6 GHz without the need to replace any parts (as is the case for conventional near-field chambers, where the probes have a limited frequency range). More information about this kind of system can be found in Section 7.6.4.

7.5 Gain Measurements

There are several methods of measuring the absolute gain of an antenna, these are discussed below.

7.5.1 Comparison with a Standard-Gain Horn

The signal obtained from the AUT is compared with that from an antenna of known gain, such as a standard-gain horn, with constant power transmitted from the source antenna. The gain of the antenna under test, G_{AUT}, is given by

$$G_{AUT} = \frac{P_{AUT}}{P_{SG}} G_{SG} \tag{7.17}$$

where P_{AUT} and P_{SG} are the powers accepted by the antenna under test and the standard gain antenna respectively (allowing for mismatches) and where G_{SG} is the known gain of the standard-gain horn. The advantage of this method is that it does not rely on knowledge of the pathloss due to the spacing between the source and test antennas in the chamber. Accuracy is determined by the positioning of the AUT and standard-gain antennas and the calibration of the standard-gain horn. It is difficult and expensive to measure the absolute gain of standard-gain horns to within a few tenths of a dB. In three-dimensional chambers, the directivity of the standard-gain horn can be measured (via radiation pattern integration) and assumed to be equal to the gain for horns with a metal-only construction (the losses are very low, so the horn can be assumed to be 100% efficient).

7.5.2 Two-Antenna Measurement

This relies on knowledge of the chamber pathloss, L, given by (see Section 3.5.1)

$$L = \left[\frac{\lambda}{4\pi l}\right]^2 \tag{7.18}$$

where λ is the wavelength and l is the separation between the source and test antennas. If the gains of the source and test antennas are denoted by G_S and G_{AUT} respectively, the power received by the latter is given by

$$P_{AUT} = L G_S P_S G_{AUT} \tag{7.19}$$

where P_S is the source antenna power. This equation is the same as Equation (3.42). If the gain of the source antenna is not known, then identical test antennas may be used at both the source and receive end of the chamber to establish their gain. Alternatively, the return signal from a single test antenna facing a large conducting sheet may be used to measure its gain (the effective chamber length being twice the separation of the antenna and sheet). In this case, it is important that the inherent mismatch of the test antenna is allowed for.

7.5.3 Three-Antenna Measurement

If three antennas of unknown gains G_1, G_2 and G_3 are used as source and test antennas in all three possible combinations, denoting the source power by P_S, the received powers are

$$\begin{aligned} P_{12} &= P_S L G_1 G_2 \\ P_{23} &= P_S L G_2 G_3 \\ P_{13} &= P_S L G_1 G_3 \end{aligned} \tag{7.20}$$

Thus, there are three equations in three unknowns so that all three gains may be established. These equations may be easily solved, giving

$$\begin{aligned} G_1 &= \frac{P_{12} P_{13}}{P_{23} P_S L} \\ G_2 &= \frac{P_{12} P_{23}}{P_{13} P_S L} \\ G_3 &= \frac{P_{23} P_{13}}{P_{12} P_S L} \end{aligned} \tag{7.21}$$

As for the two-antenna measurement, this method does not rely on the use of a standard-gain antenna, but does require knowledge of the pathloss between the source and test ends of the chamber.

7.6 Miscellaneous Topics

Antenna measurements are actually a very important and active subject. It is difficult to make accurate radiation measurements since there are many possible sources of error. In this section we are going to discuss some of the techniques and latest developments in antenna measurements.

7.6.1 Efficiency Measurements

The antenna efficiency is defined as the ratio of the directivity to the gain of an antenna and can also be determined by the radiation resistance and loss resistance, as shown in

Equation (4.40), i.e.

$$\eta_e = \frac{P_t}{P_{in}} = \frac{R_r}{R_r + R_L} \tag{7.22}$$

Thus, the efficiency can be obtained by measuring either the directivity and gain or the radiation resistance and loss resistance.

The antenna gain measurements were discussed in Section 7.5. The directivity is often computed based on the integration of the power pattern or based on the half-power beamwidth, as discussed in Section 4.1.2. This directivity/gain method is difficult to realize with precision in practice, since the full 3D radiation pattern is required and the estimated directivity is not accurate enough.

The Wheeler cap method, which was originally proposed by Wheeler [17], is a popular alternative method. It measures the radiation and loss resistances of the antenna. The idea is to place the AUT inside a small nonresonant metal shield (cap) to prevent its radiation, which is large enough not to disturb the current on the antenna. Thus, the real part of the measured input impedance should be the loss resistance of the antenna. The real part of the antenna impedance (radiation resistance and loss resistance) can be measured in an anechoic chamber or OATS. Once the loss resistance and the real part of the input impedance are known, we can use Equation (7.22) to calculate the efficiency. There are no specific limits on the cap size although the originally recommended radius of the cap should be around $\lambda vy\pi$, but the AUT size is limited approximately to $\lambda/4$ [18]. This has been widely used for mobile antenna efficiency measurement [19].

7.6.2 Reverberation Chambers

Reverberation chambers, which are electrically large metallic chambers with stirrers to generate a random field, have been used for EMC measurements for many years (for radiated emission and immunity tests) [20] and, from around the year 2000, have also been applied to the measurement of mobile phones. The advantage of this method is that the chamber can be relatively small and cheap (an absorber is not required) and a large field strength can be generated with limited power. The principle of reverberation chambers is illustrated in Figure 7.15.

The chamber itself is essentially a metal box with a source antenna that is used to set up a field distribution within it. An antenna under test (AUT) is placed in an approximately central location. Mode stirrers – large metallic plates or paddles that are articulated by motors – are used to randomize the field received by the AUT, i.e. it is as if waves arrive with equal probability from all angles. Both polarizations are also equally likely and the field is uniform on average away from the chamber walls. This is a reasonable approximation of the multipath field that is typically experienced by a mobile phone. Hence, it can be used to measure antenna diversity gain and characterize multiple-in and multiple-out (MIMO) antennas. Since it is a shielded environment (like the Wheeler cap), antenna efficiency can also be measured using a reverberation chamber. Some commercial products are already on the market [20].

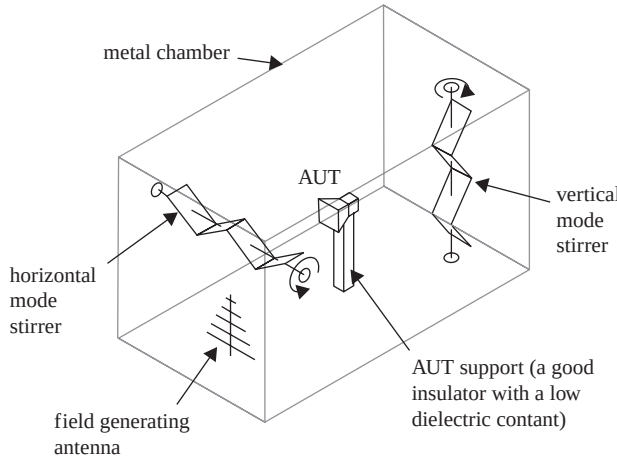


Figure 7.15 Typical reverberation chamber

7.6.3 Impedance De-embedding Techniques

Sometimes direct measurement of the impedance of an antenna may be a problem. Now let us have another look at the dual-band PIFA antenna in Example 6.4, as shown in Figure 6.26. The feed pin is right at the edge of the PCB board. It is not possible to solder a coaxial connector at this point. For measurement purposes we have to introduce a short microstrip line, as shown in Figure 6.31 and Figure 7.16. This feed line has been carefully chosen (not to disturb the antenna current distribution) to minimize its effects on the antenna performance. Although the introduction of this additional line may not affect the VSWR and return loss if there is

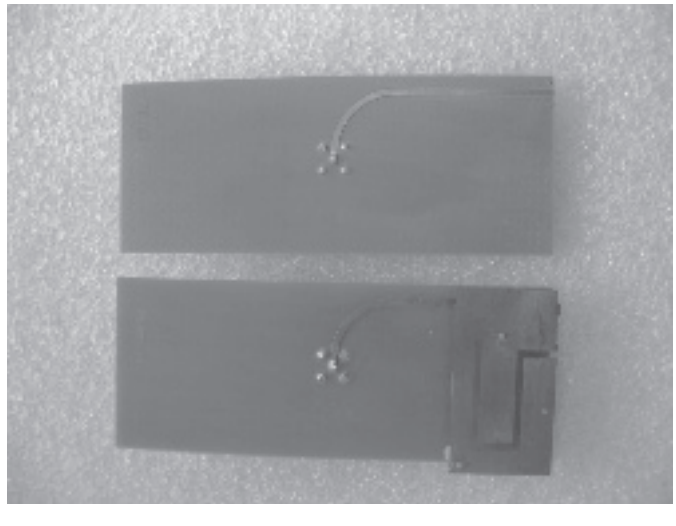


Figure 7.16 A microstrip feed line and a PIFA antenna

no mismatch and loss, it will certainly change the reading of the impedance. One method of eliminating the effects of the line and obtaining the impedance at a desired reference point is the *impedance de-embedding technique*.

The basic idea of impedance de-embedding is to make an identical feed line with an open-circuit load and then measure its reflection coefficient, which will be used to calculate the antenna impedance at the desired reference point.

The reflection coefficient at the open circuit is +1. If the length of the feed line is l, we can use Equation (2.38) to obtain the reflection coefficient at the connector of the open-circuit line:

$$S_{11_{open}} = \Gamma(l)_{open} = e^{-2\gamma l} \tag{7.23}$$

Similarly, the reflection coefficient at the connector of the antenna can be expressed as

$$S_{11_{ant}} = \Gamma(l)_{ant} = \Gamma_0 e^{-2\gamma l} \tag{7.24}$$

Thus, the reflection coefficient of the antenna (Γ_0) is the ratio of the measured S_{11ant} to S_{11open}

$$\Gamma_0 = \frac{S_{11_{ant}}}{S_{11_{open}}} \tag{7.25}$$

This is the principle of impedance de-embedding. This result can also be used to obtain the return loss, VSWR and input impedance, as discussed in Chapter 2. A formula can be yielded from Equation (2.28) to calculate the impedance:

$$Z_a = Z_0 \frac{1 + \Gamma_0}{1 - \Gamma_0} \tag{7.26}$$

As seen in Figure 7.16, an open-circuit stripline is made to match the feed line of the PIFA antenna. The measured and simulated results are given in Figures 6.32 and 6.33. Good agreement has been obtained. Without using impedance de-embedding, these results would be different, especially the impedance.

7.6.4 Probe Array in Near-Field Systems

Lars Jacob Foged
Engineering Director, SATIMO

7.6.4.1 Introduction

Further to the discussion on the near-field chamber in Section 7.4.5, we are going to present a more detailed introduction and examination on the near-field measurement techniques with reference to SATIMO systems in this section.

Spherical near-field measurements are traditionally conducted by sampling two orthogonal field components on a sphere surrounding the AUT. At any point on the spherical measurement surface the measurement antenna (probe) must point to the center of the sphere. In principle, it does not matter which of the two antennas moves relative to the other: the AUT may be

fixed, with all rotations being done by the probe; the AUT may rotate around two axes with the probe rotating around its bore-sight axis; or the AUT may rotate around one axis with the probe rotating around two axes. If the probe is dual-polarized, there is no need to rotate the probe around its bore-sight axis.

Spherical near-field antenna testing exploits the fact that the AUT can be characterized by a finite, discrete set of coefficients, which expresses the radiation and, due to reciprocity, the receiving properties of the antenna. These coefficients are the weight factors in a truncated expansion of the AUT radiation in spherical vector waves. This expansion satisfies Maxwell's equations and allows the complete characterization of the AUT radiation everywhere outside a sphere centered in the reference coordinate system and completely enclosing the AUT. The calculation of the far-field radiation from the spherical wave expansion is often referred to as *near field to far field transformation*, which was discussed in Section 5.2.1 and [22].

7.6.4.2 Traditional Systems

The actual implementation of the mechanical system can be done in a multitude of ways, with varying degrees of mechanical complexity, ranging from a fairly simple roll-over-azimuth set-up to complex double-gantry arm systems, as seen in Figure 7.17. Systems employing probes on telescopic and/or robotic arms have also been implemented. The advantage of such systems is their flexibility and also the ability to accommodate planar and cylindrical scanning geometries.

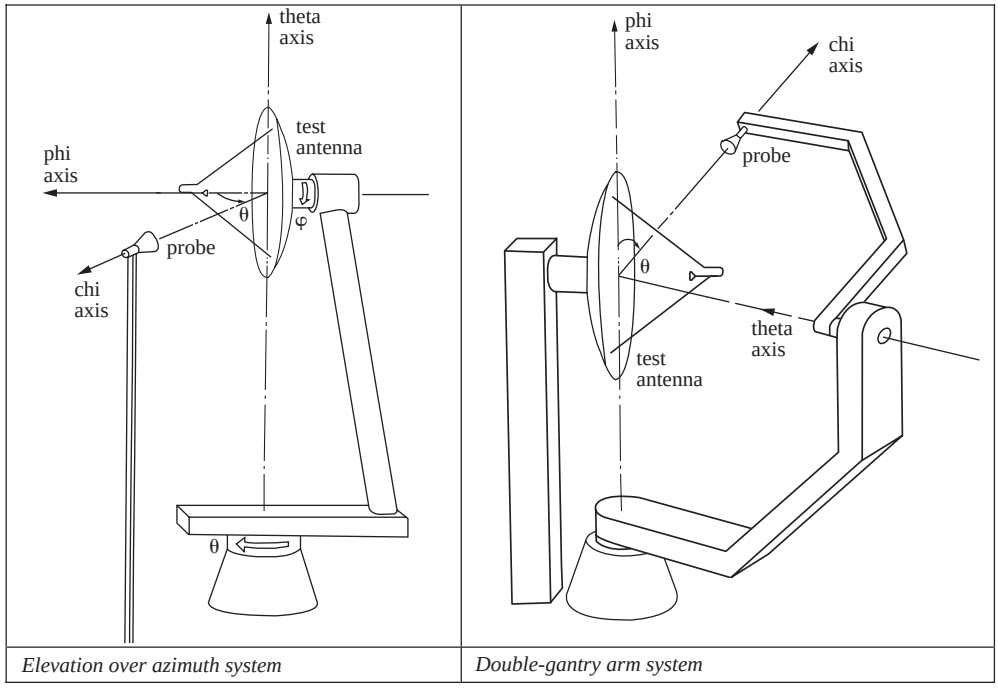


Figure 7.17 Different mechanical implementation of a spherical near-field antenna measurements system. Double-gantry arm and elevation-over-azimuth systems. (J. E. Hansen (ed.), Spherical Near-Field Antenna Measurements, IEE Electromagnetic wave series 26, Peter Peregrinus Ltd, 1988. Reproduced by permission of The Institution of Engineering & Technology (IET))

7.6.4.3 Multi-Probe Systems

In the traditional approach, described above, the mechanical movement of the probe and/or AUT is the time-consuming task with respect to the overall antenna measurement time. It is evident that an elegant way to reduce the number of mechanical movements, often two or three in traditional systems, to just one is to use several antennas in the form of an array to perform the field sampling in one dimension. This leads to a reduction in mechanical complexity, which significantly reduces the time required to complete a measurement close to a factor equal to the number of antennas in the array. An implementation of a spherical near-field system based on probe arrays is shown in Figure 7.14. The array elements are mounted on a circular arch and embedded in multilayer conformal absorbers. The probe tips protrude through small crossed slots in the smooth curvature of the absorbers, keeping the reflectivity of the probe array at a minimum. The absorbing material also reduces scattering and reflections from the support structure and cabling. The full sphere measurement is performed by electronically scanning the probe array in elevation and rotating the AUT 180 degrees in azimuth, as illustrated in Figure 7.18.

The SATIMO measurement system is based on a technique that identifies the signal from each probe in the array by perturbing the electromagnetic properties of the probe. The result of this perturbation is a modulation frequency component in the output signal, which is directly related to the amplitude and phase of the incident field at the location of the probe. By sequentially modulating each probe in the probe array it is possible to measure virtually in real time the amplitude and phase at each probe location without the need for an expensive high-frequency multiplexing network. The array of sensors is connected to the receiving equipment through a simple passive power-dividing network, as seen in Figure 7.18. This patented measurement concept is referred to as the Advanced Modulated Scattering Technique (A-MST).

Satimo's A-MST probes are dual polarized and operate over more than a decade of bandwidth. The low-frequency modulator operates in the kHz frequency range, enabling the use of standard low-cost electronic components. The low-frequency modulator is synchronized with the receiver data acquisition to provide phase and amplitude measurements at each probe in fractions of a millisecond.

The major advantage of the A-MST technology over mechanically scanned systems is the drastic improvement in measurement speed. For tests with up to a few tens of frequencies, the test time is generally reduced proportionally to the number of elements in the probe array. As

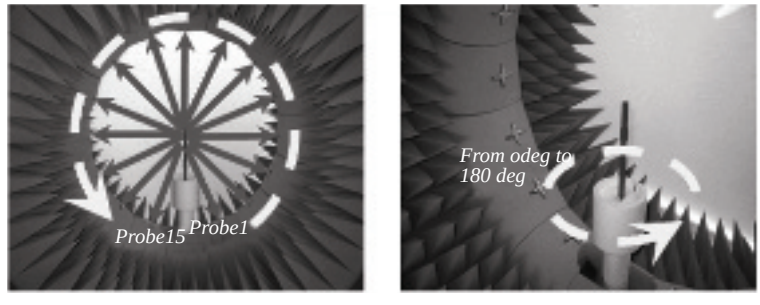


Figure 7.18 Electronic scanning for elevation sampling and AUT rotation for azimuth sampling. [Reproduced by permission of SATIMO]

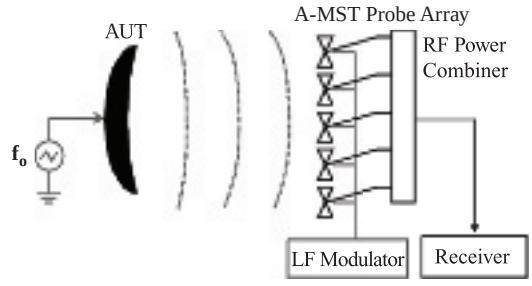


Figure 7.19 Conceptual block diagram of A-MST probe array near-field antenna measurement system. [Reproduced by permission of SATIMO]

an example, a recently installed vehicle-based antenna test system is able to do multifrequency spherical near-field measurements with 1° elevation and azimuth resolution in less than five minutes. Comparable systems based on single-probe mechanical scanning would require at least four hours for completing the same test.

7.6.4.4 Probe Array Calibration

Since the array probes have low directivity throughout the operating frequencies and the measurement distance to the AUT is not too close, the probes may be assumed to behave like Hertzian dipoles and the received signal will be proportional to the electric field parallel to the polarization of each probe. This means that traditional probe correction is not required in the near to far field transformation. However, since the probes are manufactured using high-volume industrial production techniques, the probe response cannot be assumed to be equal in amplitude, phase and polarization orientation. To correct for these discrepancies, a probe array calibration is performed.

SATIMO's probe arrays are calibrated by pointing a linearly polarized antenna toward each of the probes, as shown in Figure 7.20. Dual-polarized, phase and amplitude measurements are taken as the antenna is rotated in polarization in front of each probe. From these data a set of calibration coefficients is derived for each probe. The calibrated horizontal and vertical components of the electromagnetic field are obtained from the measurements. The process is able to compensate for differences among probes, align the polarization axes and reduce the cross-polarization components. After calibration, the typically measured probe uniformity is better than ± 0.1 dB in amplitude and $\pm 1^\circ$ in phase.

7.6.4.5 Gain Calibration

The directivity of an AUT is determined by a spherical near-field measurement followed by a near to far field transformation. During the transformation, the total radiated power is calculated and the field is normalized to obtain the directivity of the AUT in dBi.

The gain of the AUT is determined by the *gain-transfer technique*, also known as the *gain-substitution technique*, which has been adapted for spherical near-field measurement [22, 23, 24]. The gain-substitution technique requires a reference antenna with a known gain. Any

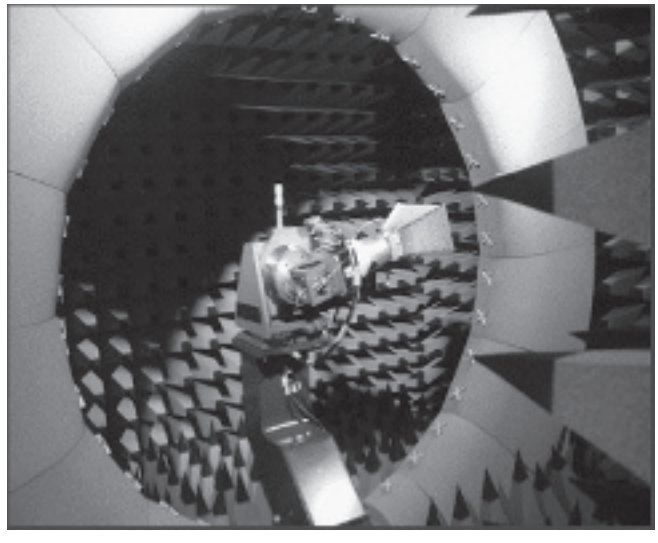


Figure 7.20 Probe array calibration using a lineary polarized antenna. [Reproduced by permission of SATIMO]

calibrated antenna can be used as reference but very often a standard-gain horn is used for simplicity. The far field of the AUT is calculated from the spherical near-field measurement through the application of a near to far field transformation. Another full-sphere measurement of a reference antenna with a well-known gain is carried out at the same frequency and with the same set-up as for the AUT and the corresponding far field of the reference antenna is calculated. Comparing the far fields of the AUT and the reference antenna, the AUT gain can be calculated.

In spherical near-field measurements there are two practical approaches for gain determination based on the known gain or efficiency performance of the reference. The two methods are often referred to as *gain calibration* and *efficiency calibration* techniques. Both approaches can determine the gain of the AUT according to the IEEE definition [12, 24].

7.6.4.6 Performance Validation

A very efficient way to validate an antenna test range is to measure a known reference antenna or compare measurements with a known reference facility. Alternatively, measurements on a reference antenna in which calculated performances can be determined with a high confidence level can be used. A double-ridged horn has been used to compare the results from different institutes using different facilities. A very good agreement has been obtained [25, 26].

7.6.4.7 Application

The probe array technology is used extensively in many demanding measurement tasks. There is no theoretical upper limit for the frequency range, although spherical near-field systems rarely go beyond 60 GHz, since very high-frequency antennas tend to be directive due to their

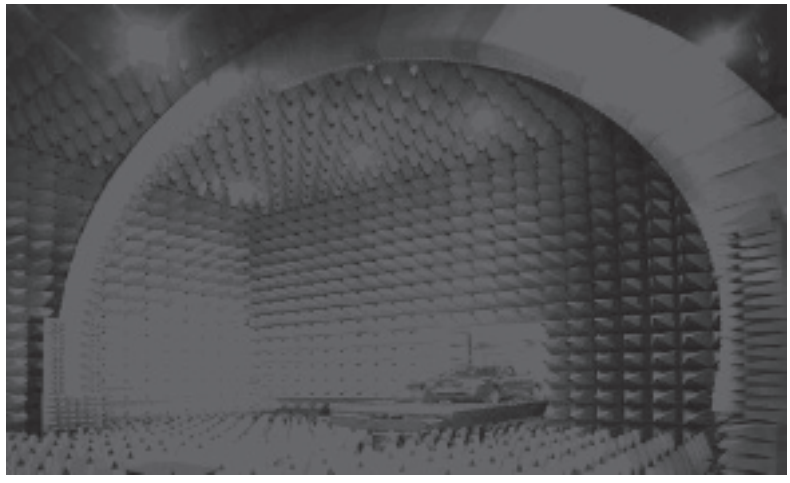


Figure 7.21 Application of probe array systems to automotive testing. [Reproduced by permission of SATIMO]

physical size and thus near-field systems based on planar scanning are often preferred due to the simpler mechanics.

The lower frequency limit for probe array technology is often imposed by the physical size of the range. Probe arrays have been used down to 70 MHz in automotive testing. Figure 7.21 shows the implementation of an automotive system for the characterization of complete car systems from 70 MHz to 18 GHz.

7.7 Summary

In this chapter we have addressed two practical issues: antenna manufacturing and measurements. The focus has been placed on the measurements. We have introduced the S-parameters and equipment and provided a good overview of the possible measurement systems with an in-depth example at the end. Some measurement techniques and problems have also been presented.

References

[1] D. Kajfez and P. Guillon (Eds), *Dielectric Resonator*, Artech House, 1986.
[2] A. A. Kishk, Y. Tin and A. W. Glisson, 'Conical dielectric resonator antennas for wideband applications,' *IEEE Trans. Antennas Propagation*, **50**, 469–474, 2002.
[3] K. M. Luk and K. W. Leung (Eds.), *Dielectric Resonator Antennas*, Research Studies Press Ltd, 2003.
[4] G. DeJean, R. Bairavasubramanian, D. Thompson, G. Ponchak, M. Tentzeris and J. Papapolymerou, 'Liquid Crystal Polymer (LCP): a new organic material for the development of multilayer dual-frequency/dual-polarization flexible antenna arrays,' *IEEE Ant. Wireless Propagation Lett*, **4**, 22–26, 2005.
[5] A. Syed, K. Demarest and D. D. Deavours, 'Effects of Antenna Material on the Performance of UHF RFID Tags,' *IEEE Int. Conf. on RFID*, 2007.
[6] H. Mosallaei and K. Sarabandi, 'Magneto-dielectrics in electromagnetics: Concept and applications,' *IEEE Trans. Antennas Propagation*, **52**(6), 1558–1567, 2004.

[7] Y. Zhang, *et al.*, 'Planar artificial magnetic conductors and patch antennas,' *IEEE Trans. Antennas Propagation*, **51**(10), 2704–2712, 2003.
[8] A. P. Feresidis, *et al*. 'Artificial magnetic conductor surfaces and their application to low-profile high-gain planar antennas,' *IEEE Trans. Antennas Propagation*, **53**(1), 209–215, 2005.
[9] Agilent: *TRL Calibration Guide*, Agilent Part Number: 85052–90059, 2005, http://cp.literature.agilent.com/litweb/pdf/85052-90059.pdf
[10] J. S. Hollis, T. J. Lyno and L Clayton, *Microwave Antenna Measurements*, Scientific-Atlanta, July 1970.
[11] W. H. Kummer and E. S. Gillespie, 'Antenna measurements – 1978', *Proc. IEEE*, **66**(4), 483–507, 1978.
[12] *IEEE Standard Test Procedures for Antennas*, IEEE Std 149-1979, published by IEEE Inc., distributed by John Wiley & Sons.
[13] G. E. Evans, *Antenna Measurement Techniques*, Artech House, 1990.
[14] Y. Huang, K. Chan and B. Cheeseman, 'Comparison of Antenna Measurements,' *Proc. European Conf. Antennas and Propagation: EuCAP 2006*, Nice, November 2006.
[15] Diamond Engineering: http://www.diamondeng.net/
[16] C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd edition, John Wiley & Sons, Inc., 1997.
[17] H. A. Wheeler, 'The radiansphere around a small antenna,' *Proc. IRE*, **47**, 1325–1331, 1959.
[18] E. H. Newman, P. Bohley and C. H. Walter, 'Two methods for the measurments of antenna efficiency,' *IEEE Trans. on AP*, **23**, 457–461, 1975.
[19] R. H. Johnston and J. G. McRory, 'An Improved Small Antenna Radiation-Efficiency Measurement Method,' *IEEE AP-Magazine*, **40**(5), October 1998.
[20] M. L. Crawford and G. H. Koepke, 'Design, Evaluation, and Use of a Reverberation Chamber for Performing Electromagnetic Susceptibility/Vulnerability Measurements,' *NBS Technical Note 1092*, National Bureau of Standards, Boulder, CO, April, 1986.
[21] http://www.bluetest.se/
[22] J. E. Hansen (Ed.), *Spherical Near-Field Antenna Measurements*, IEE Electromagnetic Waves series 26, Peter Peregrinus Ltd, 1988.
[23] P. O. Iversen, P. Garreau, K. Englund, E. Pasalic, O. Edvardsson and G. Engblom, 'Real Time Spherical Near Field Antenna Test Range for Wireless Applications,' *AMTA 1999*, California, USA.
[24] IEEE Std 145-1993, *IEEE Standard Definitions of Terms for Antennas*.
[25] L. J. Foged, Ph. Garreau, O. Breinbjerg, S. Pivnenko, M. Castañer and J. Zackrisson, 'Facility comparison and evaluation using dual ridge horn,' *AMTA 2005*, Newport RI, USA.
[26] L. J. Foged, B. Bencivenga, O. Breinbjerg, S. Pivnenko, G. Di Massa and M. Sierra-Castañer, 'Measurement facility comparisons within the European Centre of Excellence,' *AMTA 2007*, St. Louis, Missouri, USA.

Problems

Q7.1 Explain which equipment is required to measure antenna input impedance and VSWR.
Q7.2 Can we perform antenna impedance measurements inside a standard office? Justify your answer.
Q7.3 How do you measure the radiation pattern of an antenna?
Q7.4 How do you measure the gain of an antenna?
Q7.5 Explain the impedance de-embedding technique.
Q7.6 What is a network analyzer? What are its major functions for antenna measurements?
Q7.7 Explain the concept of S-parameters.

8

Special Topics

In this chapter we will address a few 'hot' topics and timely issues in antennas, which include electrically small antennas, mobile antennas, antenna diversity, multiband and ultra-wideband (UWB) antennas, radio frequency identification (RFID) antennas and reconfigurable antennas. The aim is to broaden your knowledge and provide you with information on the latest developments in antennas.

8.1 Electrically Small Antennas

8.1.1 The Basics and Impedance Bandwidth

8.1.1.1 Introduction

Whether it is to reduce wind loading, to allow more attractive products or simply to reduce costs, antenna engineers are routinely asked to reduce the size of designs without significantly compromising their performance. Hence, it is important to know that there are fundamental limits that determine the relationship between antenna size, bandwidth and efficiency. When the antenna size becomes too small, either the bandwidth or the efficiency *must* be compromised: this section gives the fundamental relations between antenna volume, bandwidth and efficiency.

The bandwidth of an antenna can be broadened to some degree using passive circuitry. However, once again, there are limits to the improvements that can be realized and the circuits introduce losses of their own. Again, this section quantifies the improvements that can be achieved.

Though somewhat arbitrary, an antenna is often considered to be electrically small when the following condition is met

$$\frac{r}{\lambda} \leq \frac{1}{2\pi} \tag{8.1}$$

where r is the radius of a sphere that just contains the antenna and λ is the wavelength. The concept of a sphere that just encloses the antenna is illustrated in Figure 8.1. Note that the feed terminals are not necessarily at the center of the sphere. We will see that this concept is important to the theory and interpretation of minimal antenna Q.

Antennas: From Theory to Practice Yi Huang and Kevin Boyle
© 2008 John Wiley & Sons, Ltd

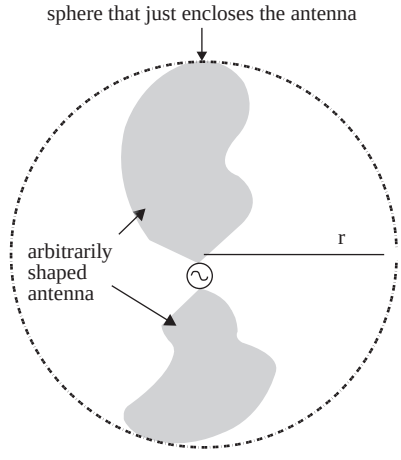


Figure 8.1 A sphere that just encloses the antenna

In the remainder of this section on the basics of small antennas, first *slope parameters* are introduced and related to the simple series and parallel resonant circuits of Chapter 2 (where the quality factor, Q, is derived). It is shown that the quality factor can be expressed in terms of the resistance and the derivative of the reactance with frequency. This applies at any frequency, not just at resonance, allowing the basic antenna Q to be determined from the antenna impedance (and variation with frequency), independent of any matching networks.

Secondly, relations between the quality factor and fractional impedance bandwidth are derived, with particular emphasis given to the criteria for maximum bandwidth.

Thirdly, the fundamental limits of antenna size, the unloaded quality factor and efficiency are reviewed.

Finally, the limits of bandwidth broadening using passive circuits are summarized. This yields the maximum possible bandwidth available from an antenna of a given size that is matched with an idealized circuit. This circuit requires an infinite number of components. Hence, a more practical analysis is given, focusing particularly on single-stage matching circuits, often referred to as *double-tuning circuits* [1]. This analysis is extended to include the case where the bandwidth-broadening components have finite losses. A simple, but important, relation is derived showing the efficiency of a double-tuning circuit as a function of the antenna and circuit quality factors. This relation puts the concentration on double tuning in context, indicating that higher order bandwidth broadening (via additional circuitry) is often not worthwhile.

8.1.1.2 Slope Parameters

Series and parallel slope parameters, x and b respectively, are defined [2] as

$$x = \frac{\omega}{2}\frac{dX}{d\omega} \tag{8.2}$$

$$b = \frac{\omega}{2}\frac{dB}{d\omega} \tag{8.3}$$

where X and B are the reactance and susceptance respectively. These parameters allow distributed circuits such as transmission lines to be represented by series or parallel lumped equivalent circuits.

The slope is normally specified only at resonance or anti-resonance (for series and parallel equivalent circuits respectively). However, here we allow the slope to be calculated at any frequency and show that the use of the slope parameters is quite general.

Using the series resonant circuit of Figure 2.21, the series slope parameter, x, is given by

$$x = \frac{1}{2}\left(\omega L + {}^{1}\!/\!_{\omega C}\right) \tag{8.4}$$

Comparing this with Equation (2.57), the generalized formula for the quality factor, gives $x = QR$. Thus, the quality factor can immediately be written in terms of the slope parameter

$$Q = \frac{x}{R} = \frac{\omega}{2R}\frac{dX}{d\omega} \tag{8.5}$$

It is not necessary to insist that the circuit is resonant, as is often assumed. It is convenient to apply this description to antenna problems, where the antenna has a series resonant nature but is not resonant: in particular, small antennas often have such characteristics. Similarly, the parallel slope parameter can be applied to antennas with substantially parallel resonances. In the parallel case

$$Q = \frac{b}{G} = \frac{\omega}{2G}\frac{dB}{d\omega} \tag{8.6}$$

where G is the conductance of the circuit.

So far the Q factor has been defined for series and parallel circuits, both in a fundamental form and in a slope parameter form that is readily applied to distributed components and some antennas. It has also been shown in Chapter 2 that the unloaded Q and fractional bandwidth only obey the commonly used relation $B_F = 1/Q$ for series or parallel circuits. In the text that follows, slope parameters are used for antennas that display predominantly series resonant impedance characteristics. In such cases, Equation (8.5) is used to relate the slope parameters to the quality factor. However, it should be understood that the quality factor, when derived in this way, is approximate, with an accuracy dependent on how well the antenna response can be modeled as a series resonant circuit (over a narrow bandwidth). *For antenna problems, the loaded quality factor is required in order to find the impedance (power transfer) bandwidth.*

8.1.1.3 Impedance Bandwidth

The unloaded bandwidths of simple circuits are related to their unloaded quality factors in Section 2.3.3. The bandwidths are based on power delivered to the load with assumed perfect sources (i.e. a voltage source is assumed for a series circuit and a current source for a parallel circuit). In practice, power transfer will depend on the relationship between the source and load impedances. Hence, it is important to consider the impedance bandwidth of the resonant circuit and to go on to find the bandwidth of power transfer from the source to the load. The most straightforward way to do this is to consider the bandwidth over which the reflection coefficient,

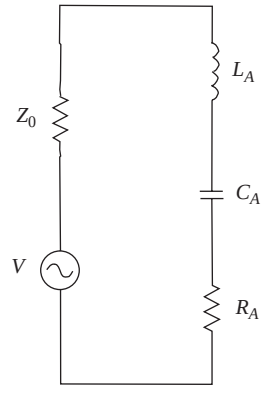


Figure 8.2 Series resonant antenna fed by a resistive source. (Reproduced by permission of Delft University Press)

ρ, or transmission coefficient, τ, is less than some predetermined limit. An example might be to consider the points at which $|\rho|^2 = |\tau|^2 = 0.5$, since this corresponds to the 3 dB points of power transmission with a reflection coefficient of 0.707. In practical designs a reflection of half the power at the antenna is seldom tolerated and the reflection coefficient is normally specified to be lower than this: for example, mobile phone antennas are often designed to a specification of $\rho = 0.5$ (still quite high, reflecting the difficulty associated with small antenna design and the variability of the antenna impedance when the phone is used), whereas HF (high-frequency) broadcast antennas (dipole arrays that operate in a number of bands over a total bandwidth of approximately an octave) are often designed for $\rho = 0.2$.

We assume that a series resonant circuit can be used to represent an antenna and, as such, give the relevant components the subscript A, as shown in Figure 8.2.

Taking $Z_0 \kappa R_A$, the antenna transmission coefficient is given by

$$|\tau|^2 = 1 - |\rho|^2 = \frac{4\kappa R_A^2}{(1+\kappa)^2 R_A^2 + X_A^2} \tag{8.7}$$

where R_A and X_A are the antenna resistance and reactance respectively, Z_0 is the source characteristic impedance and κ is a constant of proportionality. Rearranging gives

$$X_A = \pm \frac{R_A}{|\tau|^2} \sqrt{4\kappa - (1+\kappa)^2 |\tau|^2} \tag{8.8}$$

This is related to the frequency response via Equation (2.63). Hence, we need to solve

$$Q_0 \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) = \pm \frac{1}{|\tau|} \sqrt{4\kappa - (1+\kappa)^2 |\tau|^2} \tag{8.9}$$

where Q_0 is the resonant Q. This is a quadratic equation with two positive solutions for ω. The difference between these two solutions gives the fractional bandwidth:

$$B_F = \frac{\omega_2 - \omega_1}{\omega_0} = \frac{1}{Q_0}\left\{\frac{1}{|\tau|}\sqrt{4\kappa - (1+\kappa)^2\,|\tau|^2}\right\} \tag{8.10}$$

For $\kappa = 1$ (i.e. $Z_0 = R_A$), this simplifies to

$$B_F = \frac{2}{|\tau|\ Q_0}\sqrt{1 - |\tau|^2} = 2\left|\frac{\rho}{\tau}\right|\frac{1}{Q_0} \tag{8.11}$$

Taking $|\rho|^2 = 0.5$ (equivalent to the 3 dB transmission bandwidth) gives

$$B_F = 2\frac{1}{Q_0} \tag{8.12}$$

Compared with Equation (2.66), it is clear that the transmission (or loaded) bandwidth is twice the unloaded bandwidth previously calculated in Section 2.3.3. The 3 dB transmission bandwidth is equal to the inverse of the loaded Q at resonance. In turn, this is half the unloaded Q when the resistance of the load is equal to the resistance of the source.

To find out what the optimum matching impedance is for maximum bandwidth, we can find the maximum of Equation (8.10) with κ. Differentiating gives

$$\frac{d(B_F)}{d\kappa} = \frac{1}{2\,|\tau|\,Q_0}\frac{4 - 2\,(1+\kappa)\,|\tau|^2}{\sqrt{4\kappa - (1+\kappa)^2\,|\tau|^2}} \tag{8.13}$$

Setting this to zero gives

$$\kappa = \frac{2}{|\tau|^2} - 1 \tag{8.14}$$

Note that this relation is independent of Q. Since $|\tau|^2$ is always less than 1, it can be seen that $Z_0 > R_A$ is required to maximize the bandwidth. For example, if we require that $|\tau|^2$ is greater than 0.75 (equivalent to a 6 dB return loss or better), $Z_0 = 1.67R_A$.

The relations derived above can also be applied to a parallel circuit by taking $R_A = \kappa Z_0$. Here it is required that $Z_0 < R_A$.

Substituting Equation (8.14) into Equation (8.10) gives the optimal fractional bandwidth (i.e. the maximum bandwidth), B_{Fopt}:

$$B_{Fopt} = 2\frac{|\rho|}{|\tau|^2}\frac{1}{Q_0} \tag{8.15}$$

This is the maximum fractional bandwidth available from a series or parallel resonant antenna with an optimal resistance at resonance (for a given reflection coefficient). The system bandwidth and target reflection coefficient are usually known, from which the required antenna Q can be calculated. Alternatively, the system bandwidth and antenna Q may be known and it

may then be desirable to find the minimum possible reflection coefficient that can be achieved over the band.

Rewriting Equation (8.15) gives

$$B_{Fopt} = \mathbf{2}\frac{|\rho|}{1-|\rho|^2}\frac{1}{Q_0} \tag{8.16}$$

This yields a quadratic equation in $|\rho|$ which has a positive solution:

$$|\rho_{OPT}| = \frac{\sqrt{1+\left(B_{Fopt}Q_0\right)^2}-1}{B_{Fopt}Q_0}. \tag{8.17}$$

Substituting Equation (8.17) into Equation (8.14) and simplifying gives

$$Z_0 = R_A\sqrt{1+\left(Q_0 B_{Fopt}\right)^2}. \tag{8.18}$$

This gives the antenna resistance required to minimize the reflection coefficient over the required bandwidth.

Whether we optimize for bandwidth or reflection coefficient, it is necessary to know how large the antenna has to be in order to achieve a certain Q.

8.1.1.4 Fundamental Limits of Antenna Size, Q and Efficiency

It has long been recognized that, for a given frequency, the size of an antenna cannot be indefinitely reduced without compromising the Q, the efficiency or both. This problem was first explored in the 1940s. In 1947, Wheeler [3] derived a relation between antenna volume and the maximum achievable 'power factor' (equal to the inverse of the quality factor). In 1948, Chu [4] extended Wheeler's analysis by expressing radiated fields in terms of spherical modes. Chu's method is general and can be applied to any antenna; however, to allow this, Chu specified that the antenna should be contained within a sphere (as indicated in Figure 8.1) in which no energy can be stored. This is an idealized, rather optimistic, assumption that makes the underlying mathematics tractable and has often been used in subsequent work. Most antennas will store some energy within this sphere, increasing the Q. Chu used a partial fraction expansion of the wave impedance of spherical modes that exist outside the sphere bounding the antenna to obtain an equivalent (approximate) ladder network from which the Q can be calculated using circuit analysis.

A second phase of work on the limits of small antennas occurred in the 1960s. In 1960, Harrington [5] related antenna size, minimum Q and gain (including losses) for linearly and circularly polarized waves. In 1964, Collin and Rothschild [6] presented an exact method for finding the minimum Q without using the approximate equivalent network of Chu, for both spherical and cylindrical modes. In 1969, Fante [7] extended these results to include antennas with mixed polarization.

Nearly thirty years then passed before a third phase of work was undertaken in the mid-1990s. McLean [8, 9] presented a simple but exact method of determining the minimum antenna Q,

based on the observation that the lowest order TM_{01} and TE_{01} spherical modes have fields that correspond to infinitesimal electric and magnetic current elements respectively. Both linear and circular polarization can be treated this way. In 1999, Foltz and McLean [10] showed that antennas constrained to shapes other than spheres and cylinders could also be treated. More recently, Thiele *et al*. [11] addressed the question of why the previous theory has been found to be too optimistic in practice.

From all of the efforts of the workers listed above, as well as others, we know that the minimum unloaded Q at resonance of a linearly polarized antenna (when a single lowest order TM mode is implied) is given by

$$Q_0 = \eta_A \left(\frac{1}{(\beta r)^3} + \frac{1}{\beta r} \right) \tag{8.19}$$

where r is the minimum radius of a sphere that just encloses the antenna, β is the wave number $(2\pi/\lambda)$ and η_A is the antenna efficiency. It should be noted here that the antenna may be brought to resonance using an external component. Should this be the case, the Q in Equation (8.19) is that of the antenna and component combined. From Equation (8.1), for small antennas βr is less than unity and the $1/\beta r$ term is small compared to $1/(\beta r)^3$; hence, Equation (8.19) can be approximated by

$$Q_0 \approx \eta_A \left(\frac{1}{(\beta r)^3} \right) \tag{8.20}$$

Since the volume of the sphere containing the antenna is proportional to the cube of its radius, the unloaded quality factor is proportional to efficiency and inversely proportional to volume.

The fundamental limits of antenna size, Q factor and efficiency are shown graphically in Figure 8.3. For all points on and above the solid black line, the antenna can be 100% efficient. However, for all points below this line it can be seen that small antennas exhibit a trade-off between size, Q factor (bandwidth) and efficiency. Clearly there is little advantage in trading loss for bandwidth. The only other mechanism that allows these curves to be exceeded is that of interaction, either between the antenna and its supporting structure or via parasitic coupling. In the case of mobile communications equipment, both the antenna and the handset (particularly the PCB) that carries the radio are fed by the source. Radiation from the PCB increases the effective volume of the antenna and, hence, the bandwidth.

8.1.1.5 The Limits of Bandwidth Broadening

Narrowband antennas often exhibit impedance characteristics that can be modeled by series or parallel resonant circuits. It is easy to show that the bandwidth of a series resonant circuit can be enlarged using a shunt-connected parallel circuit of the same resonant frequency. Similarly, the bandwidth of a parallel resonant circuit can be enlarged by the series connection of a series resonant circuit. This is often referred to as *double tuning* [1]. For example, consider the series resonant antenna shown in Figure 8.4, with a corresponding parallel resonant double-tuning circuit formed by L_M and C_M.

It is interesting to look at the energy storage and, therefore, the Q factor of this configuration. The energy stored and dissipated in the series components is given by Equation (2.56) whereas

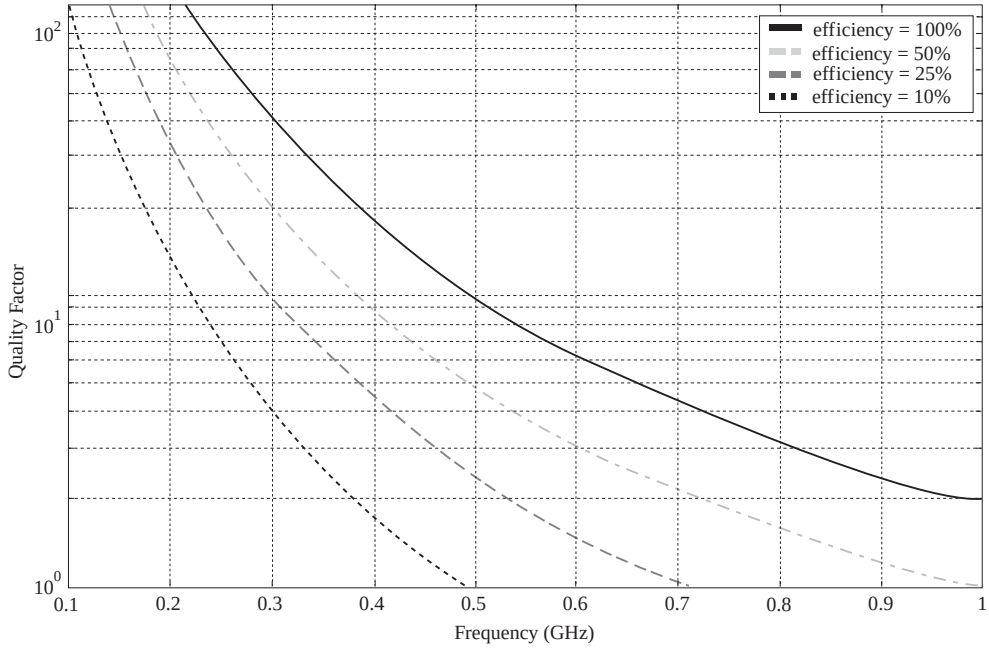


Figure 8.3 The fundamental limits of antenna size, bandwidth and efficiency. (Reproduced by permission of Delft University Press)

the energy stored in the parallel components is given by Equation (2.68). Combining these in Equation (2.52) gives the following expression for Q.

$$Q = \omega \frac{\left[\frac{1}{2} L_A I^2 + \frac{1}{2} C_A \frac{1}{(\omega C_A)^2} I^2\right] + \left\{\frac{1}{2} C_M V^2 + \frac{1}{2} L_M \frac{1}{(\omega L_M)^2} V^2\right\}}{R_A I^2} \tag{8.21}$$

Here $[\ldots]$ is used to represent the contribution to stored energy of the series reactances and $\{\ldots\}$ is used to represent the contribution of the parallel, double-tuning reactances. I is the

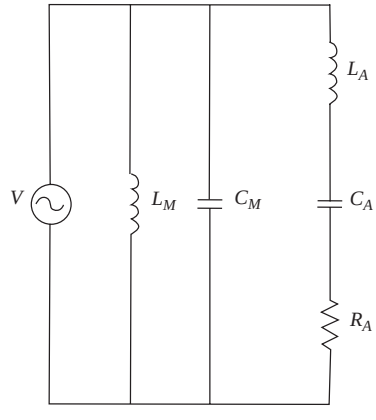


Figure 8.4 Series-parallel resonant circuit. (Reproduced by permission of Delft University Press)

current in the series branch of the circuit. This can be simplified to

$$Q = Q_A + Q_{ME} \tag{8.22}$$

Here Q_A is the Q of the series arm, while Q_{ME} is the equivalent Q of the parallel circuit, with a conductance equal to $1/R_A$. Hence, the series and parallel contributions to Q can be considered to be additive and the total Q is increased. However, if the parallel and series arms of the circuit are resonant at the same frequency, the impedance bandwidth of the circuit is increased (this will be seen later in this section). Thus, we have a seemingly contradictory situation where both the impedance bandwidth and the Q increase simultaneously. This illustrates a basic problem – there is no strict relationship between antenna Q and bandwidth. The formulas derived previously can only be applied (with care) when the antenna impedance can be closely modeled as either a series or parallel resonant circuit. Fortunately, this limitation can often be satisfied, particularly for narrow bandwidths.

Bandwidth-broadening concepts of simple, for example resonant, circuits were generalized by Bode [12] and in two well-known papers by Fano [13, 14]. Fano's theory states that the maximum achievable fractional bandwidth, $B_{F\infty}$ of a series or parallel resonant circuit in combination with an optimal bandwidth-broadening network, comprised of an infinite number of elements, is given by

$$B_{F\infty} = \frac{\pi}{Q_0 \ln\left(\frac{1}{|\rho|}\right)} \tag{8.23}$$

where $|\rho|$ is a chosen maximum reflection coefficient. The improvement in bandwidth offered by the infinite-order bandwidth-broadening network, F_∞, is given by the ratio of Equations (8.23) and (8.15), as follows:

$$F_\infty = \frac{B_{F\infty}}{B_{Fopt}} = \frac{\pi\,|\tau|^2}{2\,|\rho| \ln\left(\frac{1}{|\rho|}\right)} \tag{8.24}$$

Interpreting the theory presented is not straightforward, since these relations only strictly apply to antennas that can be represented by series or parallel equivalent circuits. Often, antenna designs that are reported as low Q are designs that incorporate lumped or distributed bandwidth-broadening circuitry on the antenna itself. The widely cited Goubau antenna [15] is believed to be an example of this. When this is the case, some of the available bandwidth broadening has already been realized on the antenna, such that only reduced improvement is possible from an external circuit.

Assuming that the antenna is well represented by a series or parallel LCR circuit, the formula given previously for the limits of (linearly polarized) antennas and bandwidth broadening can be combined to yield

$$BW_{MAX} = \frac{\pi}{\eta_A\left(\frac{1}{(\beta r)^3} + \frac{1}{\beta r}\right) \ln\left(\frac{1}{|\rho|}\right)} \tag{8.25}$$

This relation gives the maximum possible bandwidth available from an antenna of a given size that is matched with an idealized circuit containing an infinite number of components. However, it is rather optimistic, predominantly for two reasons:

- It is assumed that the antenna stores no energy within the enclosing radius, r. In practice, this is never achieved.
- Practical bandwidth-broadening circuits have finite losses – this is addressed in more detail in the text that follows.

Quite clearly, a network of infinite order is impractical. Hence, it is useful to consider how much improvement can be gained from a network with just a limited number of stages. In the following sections one stage is considered in detail – i.e. a double-tuning network. The analysis is complete, i.e. losses are included. The results of this analysis are then extrapolated qualitatively to circuits of higher order.

Consider a series resonant antenna that is double-tuned by a parallel resonant LC circuit (as shown in Figure 8.4). This operation is shown graphically in Figure 8.5.

The admittance of the antenna, Y_A, is given by

$$Y_A = G_A + jB_A = \frac{R_A - jX_A}{R_A^2 + X_A^2}. \tag{8.26}$$

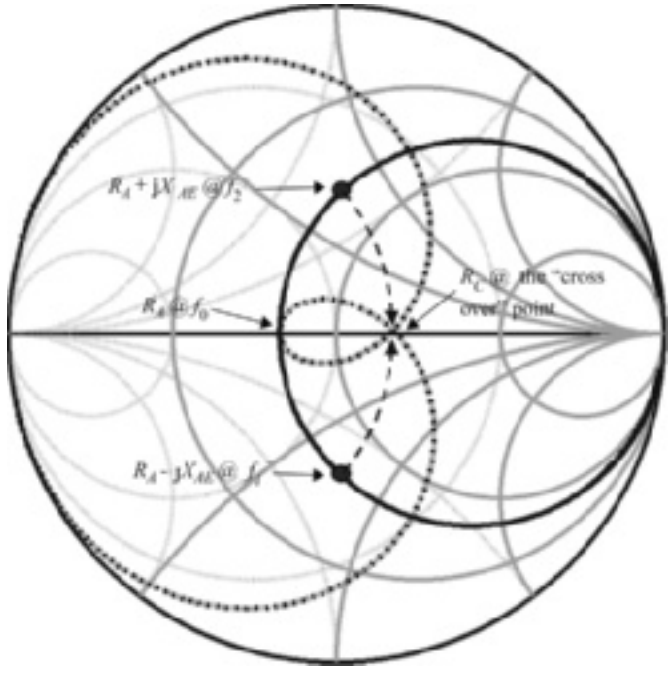


Figure 8.5 Series resonant circuit with parallel double tuning (the solid line is the S_{11} of the series resonant circuit, the dashed line is the S_{11} after double tuning, the arrows show the movement of the band edges). (Reproduced by permission of Delft University Press)

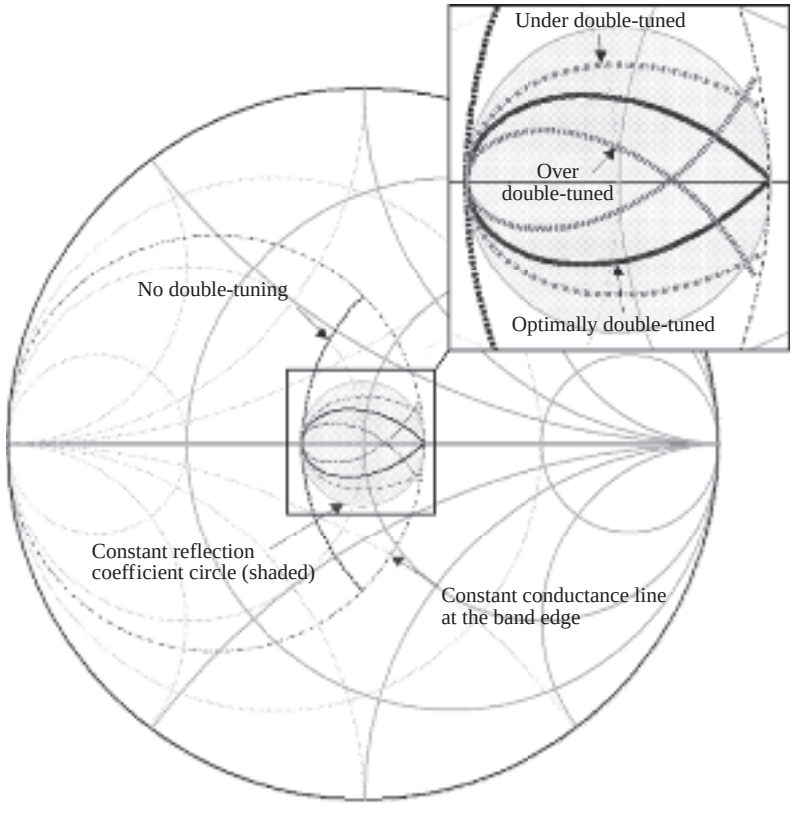


Figure 8.6 Series resonant circuit with different levels of parallel double tuning. (Reproduced by permission of Delft University Press)

Optimum double tuning occurs when the susceptance of the antenna is tuned to zero at the band edges. In Figure 8.5 this is termed the *cross-over point*. It is not worth double tuning beyond this point. Doing so would only increase the band-edge mismatch, since the constant conductance circle that passes through the band edges lies outside the constant reflection coefficient circle that passes through the point at which the antenna is resonant, as illustrated in Figure 8.6. When the antenna is optimally double-tuned it is observed that the maximum reflection coefficient is always along the resistive axis.

The resistance associated with the double tuning cross-over point, R_C , is given by

$$R_C = \frac{R_A^2 + X_{AE}^2}{R_A} \tag{8.27}$$

where X_{AE} is the reactance at the band edges (denoted by f_1 and f_2). The optimum reference impedance is then given by

$$Z_0 = \sqrt{R_A R_C} = \sqrt{R_A^2 + X_{AE}^2}. \tag{8.28}$$

Using Equation (2.63) expressed in terms of the fractional bandwidth gives

$$X_{AE} = Q_{A0} R_A \left\{ \frac{B_F (B_F + 4)}{2 (B_F + 2)} \right\} \approx Q_{A0} R_A B_F \tag{8.29}$$

For fractional bandwidths of less than 20% this approximate relation is accurate to within 5% and is therefore reasonable for most narrowband antennas. Substitution in Equation (8.28) gives

$$Z_0 = R_A \sqrt{1 + (Q_{A0} B_F)^2} \tag{8.30}$$

This relation allows the antenna resistance to be optimized (for a particular characteristic impedance) prior to double tuning. By comparing Equations (8.30) and (8.18) it can be seen that the antenna resistances required for an optimized reflection coefficient with and without double tuning are the same.

Having found the required resistance, it is then necessary to find the double-tuning components required. From Equations (8.26) and (8.28), the antenna band-edge susceptance, B_{AE}, is given by

$$B_{AE} = \frac{-j X_{AE}}{R_A^2 + X_{AE}^2} = \frac{-j \omega_0^2 L_A \{1 - (\omega/\omega_0)^2\}}{\omega Z_0^2} \tag{8.31}$$

The band-edge susceptance of the double-tuning circuit, B_{ME}, is given by

$$B_{ME} = \frac{\{1 - (\omega/\omega_0)^2\}}{-j \omega L_M} \tag{8.32}$$

where L_M is the double-tuning inductance.

For optimum double tuning, B_{AE} and B_{ME} should be equal and opposite. Equating Equations (8.31) and (8.32) and simplifying yields

$$L_M = \frac{Z_0^2}{\omega_0^2 L_A} \tag{8.33}$$

Since the double-tuning circuit has the same resonant frequency as the antenna, the required capacitance, C_M, is given by

$$C_M = \frac{1}{\omega_0^2 L_M} \tag{8.34}$$

The relations above allow the double-tuning components to be chosen based only on the characteristic impedance of the system and the antenna resonant frequency and equivalent inductance.

Having derived the necessary relations, the bandwidth improvement available from a single double-tuning network can be found (for a given reflection coefficient) and related to the maximum possible improvement given by Fano's theory.

We have already seen that, for a particular system or characteristic impedance, the antenna resistance at resonance required for double tuning is identical to that required for an optimum match without any additional circuitry. Referring to Figure 8.5, the optimum reflection coefficient prior to double tuning is worst at the band edges. Using Equation (8.28), this is given by

$$|\rho_{OPT}| = \left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| = \left|\frac{R_A + jX_{AE} - \sqrt{R_A^2 + X_{AE}^2}}{R_A + jX_{AE} + \sqrt{R_A^2 + X_{AE}^2}}\right| \tag{8.35}$$

Using Equation (8.29) and simplifying gives

$$|\rho_{OPT}| = \sqrt{\frac{\sqrt{1 + (Q_{A0}B_F)^2} - 1}{\sqrt{1 + (Q_{A0}B_F)^2} + 1}} \tag{8.36}$$

It can be shown that this is equivalent to Equation (8.17). Again referring to Figure 8.5, the worst-case reflection coefficient with double tuning is given by

$$|\rho_{DT}| = \left|\frac{R_A - \sqrt{R_A^2 + X_{AE}^2}}{R_A + \sqrt{R_A^2 + X_{AE}^2}}\right| \tag{8.37}$$

Using Equation (8.29) and simplifying gives

$$|\rho_{DT}| = |\rho_{OPT}|^2 = \frac{\sqrt{1 + (Q_{A0}B_F)^2} - 1}{\sqrt{1 + (Q_{A0}B_F)^2} + 1} \tag{8.38}$$

Hence, for a defined bandwidth, there is a very simple relation between the reflection coefficient after double tuning, $|\rho_{DT}|$, and the optimum reflection coefficient prior to double tuning, $|\rho_{OPT}|$.

Alternatively, by rearranging Equation (8.38), the bandwidth with double tuning, B_{FDT}, can be written

$$B_{FDT} = \frac{1}{Q_{A0}} \frac{2\sqrt{|\rho|}}{1 - |\rho|} \tag{8.39}$$

where $|\rho|$ is a chosen, or target, reflection coefficient magnitude. Comparing this with the optimum bandwidth available without any circuitry, given in Equation (8.15), yields the double-tuning bandwidth improvement factor, F_{DT}, as follows

$$F_{DT} = \frac{1 + |\rho|}{\sqrt{|\rho|}} \tag{8.40}$$

The variation of F_{DT} and the maximum possible improvement, F_∞, given previously in Equation (8.24), are shown in Figure 8.7.

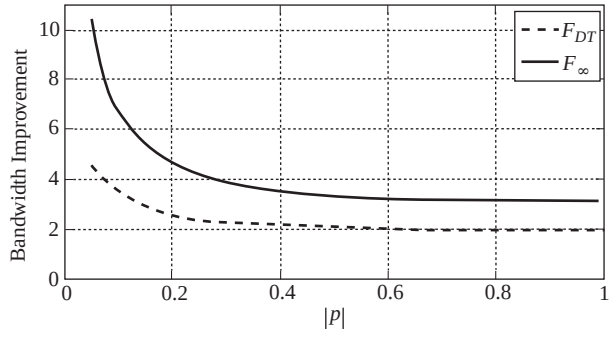


Figure 8.7 Variation of bandwidth improvement factors F_{DT} and F_∞. (Reproduced by permission of Delft University Press)

It can be seen that the improvement factors are largest for small reflection coefficients. Importantly, approximately half of the total improvement is achieved with a double-tuning network.

The analysis so far has assumed all components to be lossless. Of course, this is not the situation encountered in practice, and the effects of component losses must be considered. Here only the losses associated with double tuning will be considered. Referring to Figure 8.4, the loss of the double-tuning components can be represented by an additional parallel resistance, R_M. Using Equation (2.69), the Q of the double-tuning network, Q_M, will then be given by

$$Q_M = \frac{\left(\omega C_M + 1/\omega L_M\right)}{2G_M} = \frac{1}{\omega L_M}\left\{\frac{1+\left(\omega/\omega_0\right)^2}{2G_M}\right\} \tag{8.41}$$

We assume that the antenna resistance is optimized as defined by Equation (8.30) and that this optimum condition is unaffected by the finite Q of the double-tuning circuit (a reasonable assumption if the conductivity of the double-tuning circuit is considerably less than that of the antenna). Using Equation (8.33) gives

$$L_M = \frac{Z_0^2}{\omega_0 Q_{A0} R_A} \tag{8.42}$$

Substituting Equation (8.42) into Equation (8.41) gives

$$Q_M = \frac{Q_{A0} R_A}{2Z_0^2 G_M}\left\{\frac{\omega}{\omega_0} + \frac{\omega_0}{\omega}\right\} \tag{8.43}$$

We have previously seen that the term in $\{..\}$ is approximately equal to 2 over narrow to moderate bandwidths. Using this and rearranging for G_M gives

$$G_M = \frac{Q_{A0} R_A}{Z_0^2 Q_M} \tag{8.44}$$

Using Equations (8.26) and (8.28), the band-edge (i.e. the highest or worst case) conductance of the antenna is given by

$$G_{AE} = \frac{R_A}{Z_0^2} \tag{8.45}$$

Hence, the minimum (band-edge) efficiency, η_{DT}^{MIN}, is given by

$$\eta_{DT}^{MIN} = \frac{G_{AE}}{G_{AE} + G_M} = \frac{1}{1 + Q_{A0}/Q_M} \tag{8.46}$$

This is a very simple relation dependent only on the antenna and double-tuning circuit Q factors. There is no bandwidth dependency. Note also that Equation (8.46) does not include the effects of mismatch.

It is clear that, for good efficiency, the double-tuning circuit Q should be much higher than that of the antenna. For example, dual-band mobile phones have typical antenna Qs of approximately 15, while a lumped circuit double-tuning network will have a typical Q of, at best, 50. Using these values gives a band-edge efficiency of 77%. GSM and DCS have fractional bandwidths of approximately 10%. With double tuning, using Equation (8.38), the band-edge mismatch efficiency is 92%. This gives a combined band-edge efficiency of 71%. From Equation (8.17), the mismatch efficiency without double tuning is 71%. Hence, in this example, there is no advantage in using double tuning – the improvement in mismatch efficiency (return loss) is counter-balanced by the mismatch corrected efficiency (insertion loss) of the double-tuning circuit. Of course, there may be other reasons for using double tuning. For example, the improved antenna return loss will reduce cumulative losses in subsequent parts of the RF chain. Also, the double-tuning circuit provides a degree of filtering. However, it is clear that the use of bandwidth-broadening techniques requires the use of high Q resonant circuits (with respect to the antenna Q).

Circuits with higher orders of bandwidth broadening require resonators with increased slope parameters, increasing the losses of each stage. Also, the losses of each stage of the network will be cumulative. Hence, high-order bandwidth-broadening circuits will require very high Q resonators. Since approximately half of the maximum bandwidth improvement is provided by a single resonator with only moderate loss, higher order circuits are often not worthwhile.

It should also be recognized that personal devices such as mobile phones are used in a manner such that the driving-point impedance of the antenna is variable within limits defined by the modes of use. For example, the antenna of a mobile phone will present different impedances when in free space and when held in 'talk position'. High-order bandwidth-broadening circuits are critical and, hence, may become counterproductive when large impedance variations are experienced. Simple double-tuning circuits are thought to be more robust in this respect.

8.1.1.6 Discussions and Conclusions

Some basic and important relations that are used to determine antenna bandwidth have just been introduced. In particular, it has been shown that when antennas can be represented as

series or parallel resonant circuits, as is often the case for small antennas, the quality factor can be found from the resistance and the derivative of the reactance with frequency. This applies at any frequency, not just at resonance, and allows comparison between different designs in a fundamental way.

We have shown that antennas represented by series or parallel resonant circuits have optimum bandwidth when mismatched at resonance. Series resonant circuits are required to have a resistance at resonance that is lower than the system (or characteristic) impedance. The converse applies to parallel resonant circuits. For maximum bandwidth, the level of mismatch depends only on the reflection coefficient magnitude at which the bandwidth is measured. However, for a minimum reflection coefficient over a given bandwidth, the level of mismatch is inversely related to the antenna quality factor.

The limits of antenna size, bandwidth and efficiency have been reviewed, likewise the limits of passive bandwidth broadening. We have seen that there is no universal relationship between bandwidth and quality factor and that some care must be applied when interpreting how well a particular antenna performs. Some small antennas have distributed reactive networks integrated within the antenna structure. Such networks can provide a degree of bandwidth broadening. When this is the case, further bandwidth-broadening circuitry will give improvements lower than those stated both in this chapter and in the literature – some of the available improvement will have been 'used' on the antenna.

A combined formula has been given, showing the maximum possible bandwidth available from an antenna of a given size that is matched with an idealized circuit containing an infinite number of components.

A more practical analysis of the bandwidth broadening available from a double-tuning circuit with finite losses has been undertaken. A simple new formula for the efficiency of a double-tuning network was derived. This relation indicates the degree to which the Q of any double-tuning circuit must be higher than the Q of the antenna (or any other load) that requires broadbanding. From this result, it is reasoned that high-order bandwidth broadening is only possible if extremely good circuit technologies are used. High-order circuitry is also thought to be too critical for applications such as mobile phones, such that it could become counterproductive if the antenna impedance were to vary due to user interaction. Double tuning is thought to be less susceptible to changes in the antenna impedance that occur due to user interaction and places less stringent requirements on component quality. Double tuning can approximately double the bandwidth of an antenna designed for a reflection coefficient of 0.5 (such as a typical mobile phone antenna). This is approximately half the improvement available from a network of infinite order. Hence, often only double tuning is required – higher order bandwidth-broadening circuits are unlikely to bring significant additional improvements overall.

To maximize the benefits of bandwidth broadening, high-quality circuit technologies are required. For some applications such components are available. For example, air wound inductors and vacuum capacitors – with Q factors of around 200 and 1000 respectively – are used for AM broadcast applications. For other applications, such as mobile phones, size and technology constraints are such that inductor and capacitor Qs are typically 30 and 100 respectively. Inductors are avoided where possible. In such cases it is often more desirable to form distributed double-tuning circuits on the antenna itself.

8.1.2 Antenna Size-Reduction Techniques

There are many ways of reducing the size of an antenna. Some of the most common are:

- top loading;
- matching;
- reactive loading;
- dielectric loading.

Each of these methods will be dealt with in the sections that follow.

8.1.2.1 Top Loading

Consider the monopole above ground shown in Figure 8.8. The current distribution indicated can be maintained by keeping the length of the antenna approximately constant, but having some of the structure located parallel to the ground, as shown for the 'T', 'Star' and 'Disc' configurations. This allows the height of the antenna to be reduced without changing the resonant frequency.

However, as indicated for the 'T' antenna in Figure 8.9, currents on the upper part of the structure cancel out currents induced in the ground. Hence, while the current distribution is maintained, the part of the antenna that is parallel to the ground contributes very little to radiation. Radiation thus emanates from a smaller part of the 'T' antenna than for the monopole and the radiation resistance reduces.

However, the radiation resistance is still higher than it would be for a monopole of the same height, h, as the 'T' antenna, due to the more rectangular shape of the current distribution of the vertical part of the latter. This can be explained by considering the linear antenna shown in Figure 8.10. The current on the antenna is broken up into small elements, from which it is possible to calculate the radiation resistance for different current distributions.

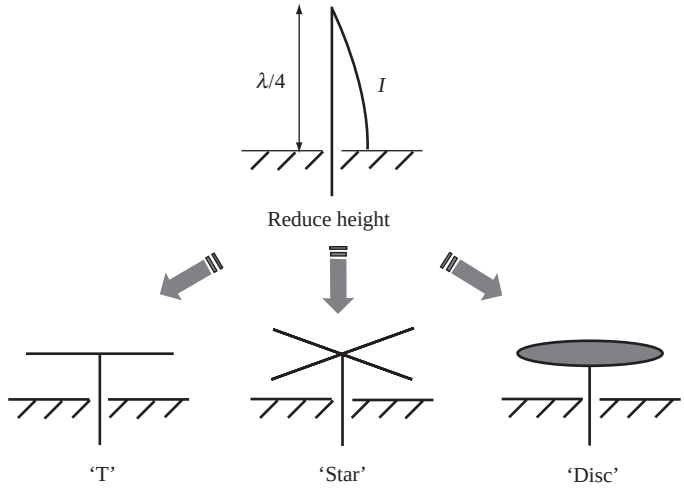


Figure 8.8 Height reduction via top loading

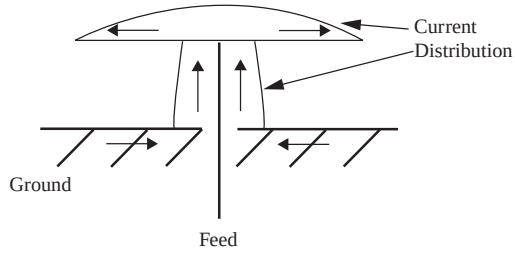


Figure 8.9 Current distribution of a top-loaded 'T' antenna

For any small linear antenna, the radiation resistance R_r, is given by

$$R_r = 80\pi^2 \left(\frac{\delta l}{\lambda}\right)^2 \left|\sum_{n=1}^{N} \frac{I_n}{I_0}\right|^2 \tag{8.47}$$

where I_0 is the maximum antenna current and I_N is the elementary antenna current. With a triangular variation of current, as occurs for very small monopoles, the radiation resistance is given by

$$R_r = 20\pi^2 \left(\frac{N\delta l}{\lambda}\right)^2 \tag{8.48}$$

For $(N\delta l)/\lambda = 0.1$, the radiation resistance is approximately equal to 2 Ω. For a top-loaded antenna of the same height, the distribution of the radiating current (the vertical part in Figure 8.9) is close to rectangular, and the radiation resistance is given by

$$R_r = 80\pi^2 \left(\frac{N\delta l}{\lambda}\right)^2 \tag{8.49}$$

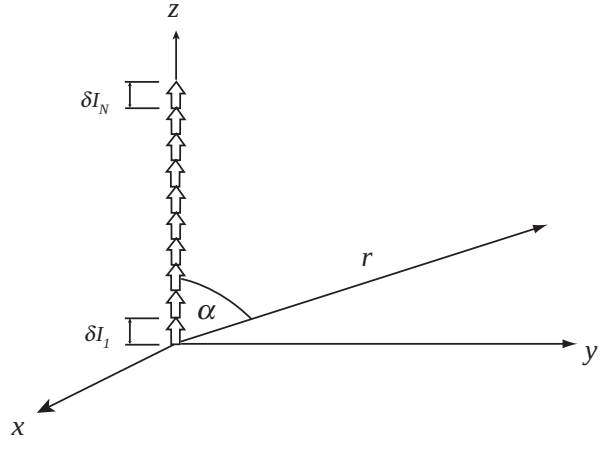


Figure 8.10 Linear antenna with elemental currents

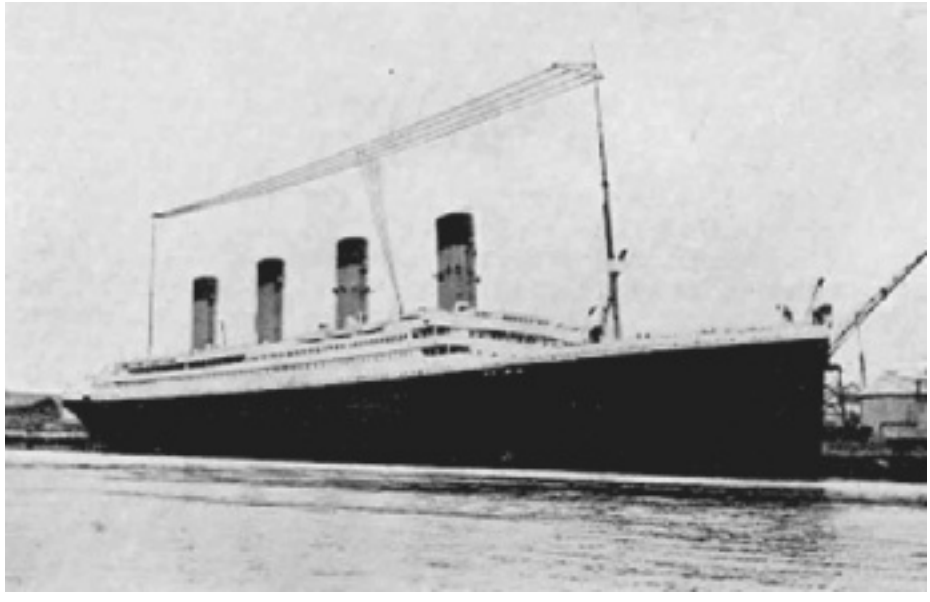


Figure 8.11 'T' on the Titanic. (Reproduced by permission of © 2001–2004 Marconi Corporation plc)

Again, taking $(N\delta l)/\lambda = 0.1$, the radiation resistance is approximately equal to 8 Ω. Inspection of Equations (8.48) and (8.49) also shows that a factor of four improvement is possible. Hence, top loading not only allows the resonant frequency to be maintained as the antenna height is reduced, it also improves the radiation resistance for a given height. Hence, it is a very popular and widespread method of achieving a low-profile structure. An early example of the use of top loading is shown in Figure 8.11.

The Titanic famously sank in 1912. Marconi-installed transmitter equipment was used to sound the alarm using the (normally not shown) large 'T' antenna.

8.1.2.2 Matching

Most electrically small antennas have a low radiation resistance and high, normally capacitive, reactance. A simple approach, illustrated in Figure 8.12, is to match this impedance to 50 Ω – or whatever the system impedance might be – using discrete components, for example two inductors, or an inductor and a capacitor.

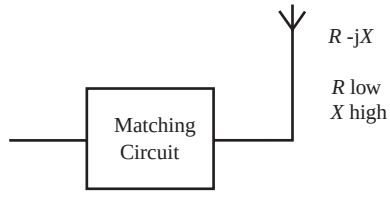


Figure 8.12 Antenna size reduction using impedance matching

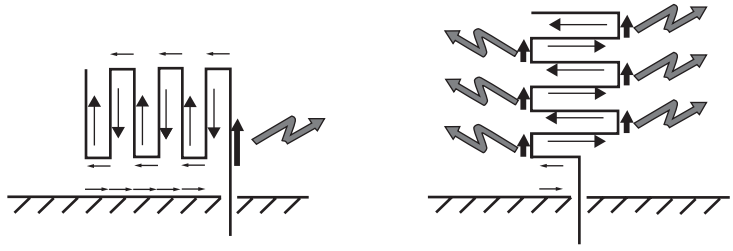


Figure 8.13 Radiation from meandered structures

Some attention must be paid to the component quality when using this technique; for example, for MF (medium frequency) broadcasting, air core inductors and vacuum capacitors are used in order to minimize losses. Litz wire – designed to have lower losses due to the skin effect than conventional solid conductors – may also be used. Many strands are used with a conductive coating on each strand. Current is spread across the strand, which increases the effective outer area of the conductor for a given circumference. The strands are also twisted in a way that minimizes the induction of opposing electromagnetic fields in other strands.

8.1.2.3 Reactive Loading

A common method of reducing the resonant frequency of a small antenna is to use some form of reactive loading. The aim is to store enough magnetic energy (to create inductance) to counter the electric energy (capacitance) that is associated with most small antennas. This is illustrated for two meandered structures in Figure 8.13.

The bold arrows show, broadly speaking, where radiation occurs. The nonbold arrows can be considered to be in pairs that have opposite directions. Hence, they cancel in the far field and do not produce any radiation. However, they do store energy in the near field, and this can be used to reduce the antenna resonant frequency.

An alternative way of viewing this is that the antenna length is maintained at approximately $\lambda/4$, so that the resonant frequency is the same as that of an equivalent $\lambda/4$ monopole. However, since radiation comes from less of the structure, the radiation resistance must be reduced. In practice, the length of the meander must be somewhat longer than $\lambda/4$ in order to achieve resonance.

Another example of reactive loading – this time applied to planar antennas – is shown in Figure 8.14. Slots and notches are cut in the antenna structure in order to lengthen the path over which current travels (energy is also stored around the slots and notches), reducing the resonant frequency.

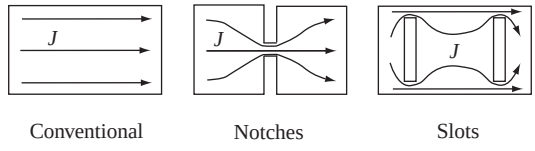


Figure 8.14 Reactive loading of planar antennas

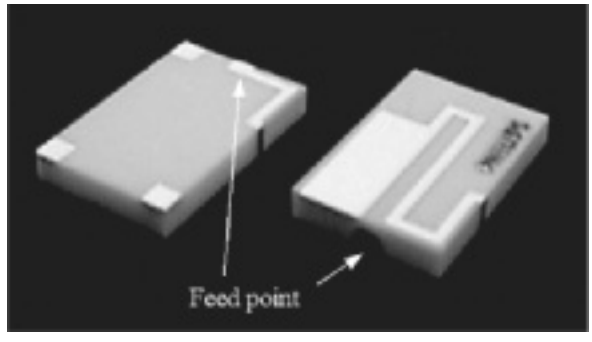


Figure 8.15 Meandered, dielectric loaded monopole antenna for dual-band mobile phones (the view on the left shows the bottom of the antenna, intended for connection to the PCB, and the view on the right shows the top of the antenna)

8.1.2.4 Dielectric Loading

A high dielectric constant material can be used to achieve a slow wave structure that allows smaller resonators to be realized. Antennas with dielectric loading have the advantage that tuning capacitance can easily be built in to the structure. For consumer wireless applications, it is also possible to excite a polarization out of the plane of a host PCB. However, dielectric antennas are not immune to the fundamental limits of bandwidth, size and efficiency: smaller means narrower bandwidth. Also, loss occurs within the dielectric material and the high Q nature of such antennas tends to increase the losses in conductors. Often these conductors have a conductivity that is compromised somewhat in order to be compatible with the dielectric material from a manufacturing standpoint.

Figure 8.15 shows a meandered, dielectric loaded monopole antenna manufactured for dual-band mobile phones. The antennas measure $17 \times 10 \times 2$ mm and the dielectric constant is approximately 20.

Figure 8.16 shows a half-wave global positioning system (GPS) patch antenna. Since the wavelength is determined by the dielectric constant, ε_r, of the patch substrate, the antenna

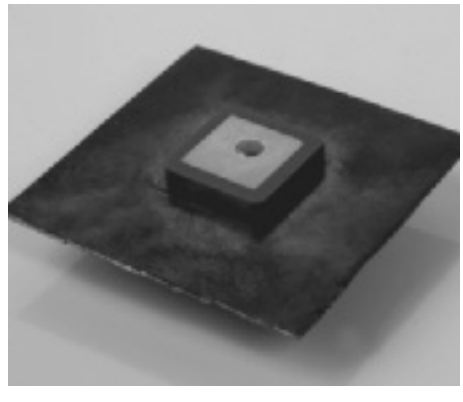


Figure 8.16 A half-wave GPS antenna mounted on a finite ground plane

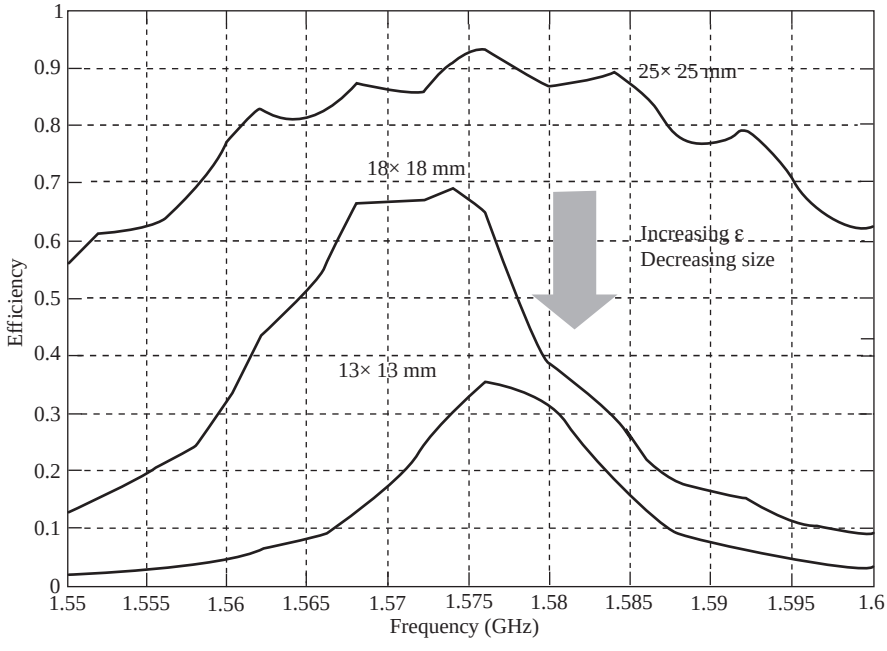


Figure 8.17 Measured efficiency of GPS patch antennas of the dimensions indicated

can be made progressively smaller by increasing ε_r. The bandwidth of GPS is very narrow, so the antenna size can be reduced significantly without necessarily reducing efficiency. However, the measured results shown in Figure 8.17 indicate that as the antenna size is reduced, there is an almost inevitable reduction in antenna efficiency (the bandwidth also reduces, as expected).

8.2 Mobile Antennas, Antenna Diversity and Human Body Effects

8.2.1 Introduction

The term *mobile radio* is rather loosely used to describe a radio communication link where at least one end of the link is either at an unknown location or in motion. Marconi's transmission from a fixed site on the Isle of Wight, England to a tug-boat located at an ill-defined position approximately 18 miles away was probably the first example of such a link in 1888. To be more specific, this was an example of maritime mobile radio. An early example of a land-based mobile transceiver is given in Figure 8.18.

Here, Marconi and Fleming are pictured with a Thornycroft steam bus outside the Haven Hotel in Poole, England just before the turn of the 20th century. Marconi used amplitude modulation (AM) at 'low frequencies', so the antenna seen attached to the roof of the bus was probably electrically small**,** as defined by Equation (8.1).

Since Marconi's first experiments, many other applications have been found which fall into the broad class of mobile radio. The first system to closely resemble the mobile radio networks

Figure 8.18 An early mobile transceiver: the antenna is located on the roof of the bus. (Reproduced by permission of © 2001–2004 Marconi Corporation plc)

that we use today was implemented by the Detroit Police Department in 1921 [16]. It used a frequency band centered at approximately 2 MHz for communication between a central controller and police cars. This was an example of a dispatch system, but other modes of communication – such as walkie-talkies, paging and mobile telephony – soon followed.

This section gives details of the design principles of modern mobile phone antennas. It also gives a summary of antenna diversity techniques – where two or more receive antennas are used rather than one – and the effect of the human body on antenna performance.

8.2.2 Mobile Antennas

8.2.2.1 The Cellular Frequency Bands

The most widely utilized European and American cellular telephony bands, which have also been adopted in many other countries, are shown in Figure 8.19.

The acronyms used are as follows:

GSM	Global System for Mobile Communications (previously Groupe Spécial Mobile)
DCS	Digital Cellular System
PCS	Personal Communications System
UMTS	Universal Mobile Telecommunications System
AWS	Advanced Wireless Service

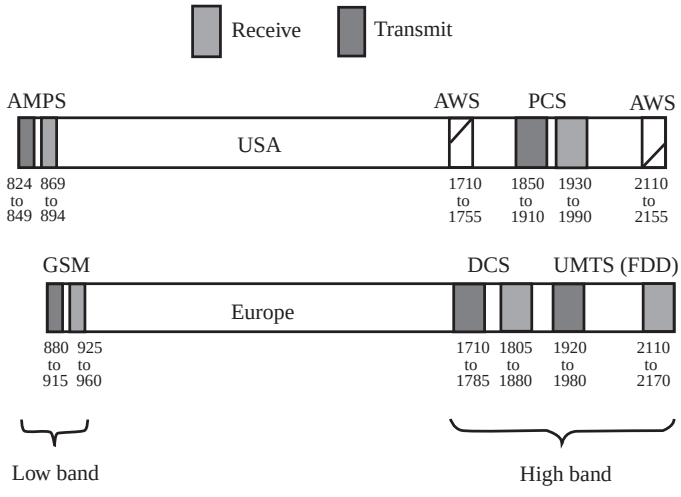


Figure 8.19 European and US cellular frequency bands (MHz). (Reproduced by permission of Delft University Press)

Initially mobile phones were designed for only one band. However, dual-band phones – for example, capable of operation in the GSM900 and DCS1800 bands – became the minimum requirement in Europe in the late 1990s. Tri-band operation in the GSM900, DCS1800 and PCS1900 bands became commonplace for 'high-end' phones in the early 2000s, allowing operation in many countries. It can be seen from Figure 8.19 that tri-band GSM/DCS/PCS antenna operation can be achieved using a fundamentally dual-band design with an extended bandwidth in the higher band. Similarly, four or five bands may also be covered by extending the bandwidths of a dual-band design even further.

In the cellular frequency bands, mobile phones transmit at relatively high power levels. GSM, DCS and UMTS have maximum average output powers of 0.25 W, 0.125 W and 0.125 W (in Class 4 operation) respectively.

8.2.2.2 The 'Connectivity' Frequency Bands

The wireless local area network (WLAN) and personal area network (PAN) bands that are most commonly used by consumers (i.e. the general public) are indicated in Figure 8.20. The Global Positioning System (GPS) band is also shown, due to its importance to location-based services. Other acronyms used are as follows:

UNII	Unlicensed National Information Infrastructure
HIPERLAN	High Performance Radio Local Area Network

In many applications these systems are only used at low power levels (for example, Bluetooth has a typical output power of 1 mW). Hence, these low-power 'connectivity' systems have very different design requirements/restrictions to cellular systems (which transmit at relatively high power), both for the antenna and for the RF circuitry.

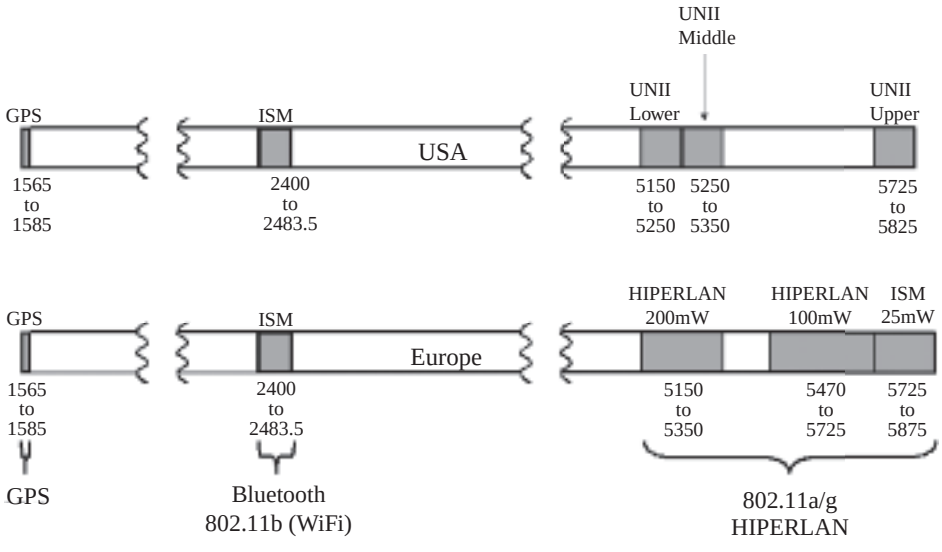


Figure 8.20 Main 'connectivity' bands (MHz)

The type of antenna that is used with a particular type of phone is normally determined by aesthetic considerations and specific absorption rate (SAR) regulations. This will be covered in more detail later in this chapter. It is a measure of the maximum amount of energy dissipated per unit volume in any part of the phone user's body. When the phone is held in 'talk position' (i.e. next to the head), this maximum often occurs close to the antenna.

There are four main types of cellular antenna:

- monopoles (whips);
- normal mode helices;
- meander line antennas (loaded/unloaded);
- planar inverted F antennas.

Monopoles and helical antennas are normally located outside the phone casing, as illustrated in Figure 8.21. There are two broad classes of internal antenna, dependent on the relative positions of the antenna and printed circuit board (PCB):

- adjacent antenna and PCB – where the antenna is installed next to the PCB (in the manner of a monopole);
- coincident antenna and PCB – where the antenna is installed on top of the PCB.

These classes are shown diagrammatically in Figure 8.22.

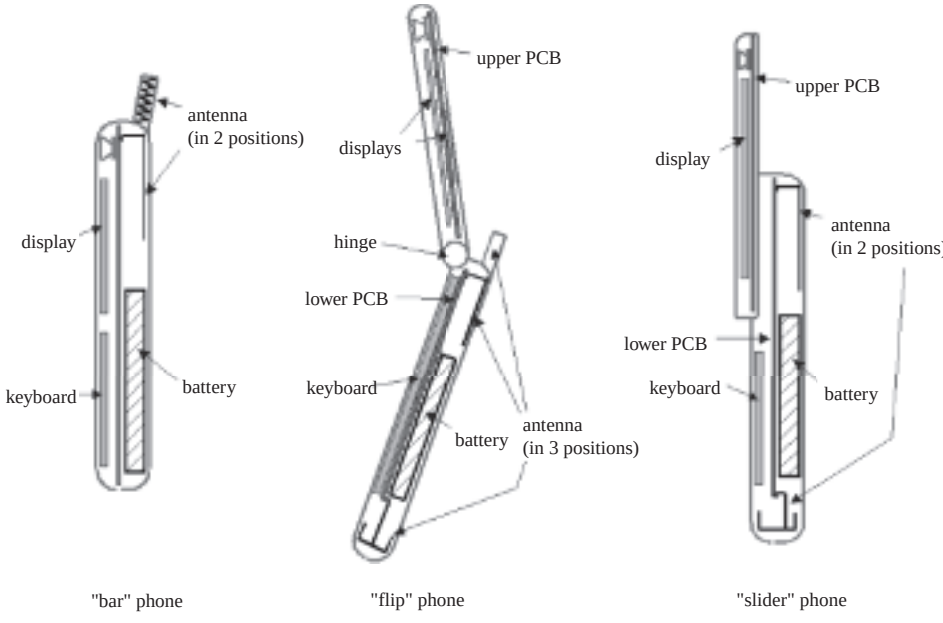
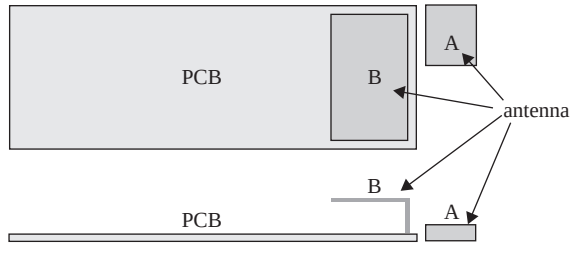


Figure 8.21 Major mobile phone types: 'bar', 'flip' and 'slider'

8.2.2.3 Typical Antenna Types

Monopoles

Monopoles were the antenna of choice for the earliest mobile phones and they are still used, particularly in countries where coverage is limited. They have the advantage of providing significant clearance between the antenna and the head, which allows low SAR and, perhaps most importantly, high efficiency to be achieved. A simple monopole located over a ground plane is illustrated in Figure 8.23.



A - Adjacent antenna and PCB
B - Coincident antenna and PCB

Figure 8.22 Basic antenna and PCB arrangements. (Reproduced by permission of Delft University Press)

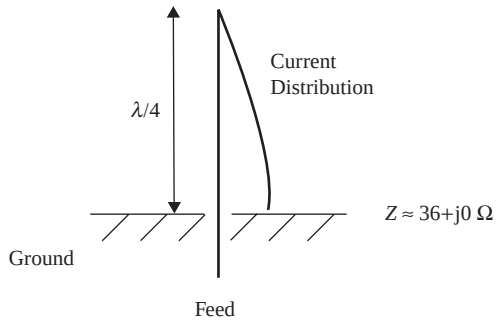


Figure 8.23 Quarter-wave monopole over a ground plane

Most monopoles are operated at an electrical length that is close to a quarter of a wavelength, when the input resistance is 36 Ω and the input reactance is zero. This impedance matches quite well with the commonly used system impedance of 50 Ω. When mounted on a typical handset, monopole antennas can be well matched to 50 Ω when slightly shorter than λ/4, due to the influence of the conductive parts of the handset, which also radiate considerably.

Helical Antennas

Helical antennas were widely used in the 1990s, as they occupied considerably less space than the monopoles that preceded them. They were an early example of a compromise between technical performance and commercial attractiveness. The performance of helical antennas is worse than that of their monopole predecessors; however, they allowed smaller, more aesthetically attractive phones to be manufactured. This compromise continued with the later introduction of internal antennas.

A typical helix antenna assembly is shown in Figure 8.24.

The performance is defined by the following parameters (refer to Figure 8.25):

C circumference
S spacing between turns
L unfolded turn length
θ pitch angle: $\tan(\theta) = S/C$

When C/λ is between 0.75 and 1.25, the antenna operates in an *axial mode*, when the radiation is along the axis of the antenna (the upward direction shown in Figure 8.24). The antenna has some useful properties in this mode: the gain is high, the polarization is circular and the impedance is close to 150 ohms over close to an octave bandwidth. Because of this, axial mode helices are widely used, for example as satellite antennas. However, they are not normally used for mobile applications.

When C/λ is less than 0.5, the helix is said to operate in a *normal mode*, when the radiation is similar to that of a monopole aligned along the axis of the helix (radiation is predominantly perpendicular to the axis and the polarization is linear, in line with the axis). In this mode the antenna is resonant with a height that is significantly less than that of a comparable monopole antenna.

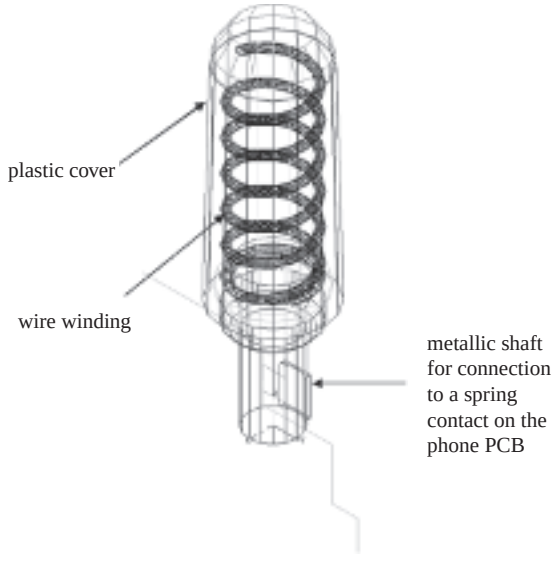


Figure 8.24 A helix antenna assembly (shown as a 'wire grid model' as displayed in a typical 3D electromagnetic simulator)

Monopole-Like Antennas

Monopole-like antennas operate in a similar fashion to helical antennas, since they are generally both reactively loaded in order to reduce size (as discussed in Section 8.1.2). The distinction made here is that helical antennas are generally external to the phone casing whereas monopole-like antennas are located internally. As such, the latter are often planar or conform to the shape of the phone.

Planar Inverted F Antennas

The planar inverted F antenna (PIFA), which was used as a major example in previous chapters, became the antenna of choice for mobile phones in the late 1990s due to the requirement to have an internal antenna with a low SAR. They can be seen as evolving from either a monopole or a half-wave patch antenna. Evolution from a monopole is illustrated in Figure 8.26.

Here a quarter-wave monopole is first folded to form an inverted L antenna (ILA), as proposed by King *et al.* in 1960 [17]. King *et al*. also proposed the second step in the evolutionary path from monopole to PIFA: that is, the introduction of a shorting pin. At the time, this configuration

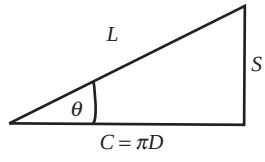


Figure 8.25 Relationship between helix dimensions

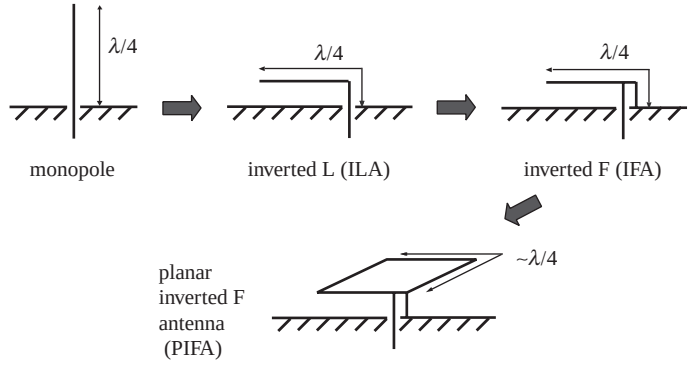


Figure 8.26 Evolution of a PIFA from a monopole antenna. (Reproduced by permission of Delft University Press)

was referred to as a *shunt-driven inverted L antenna*, though it later became almost universally known as an *inverted F antenna* (IFA) [18]. The introduction of the shorting pin allows the low impedance of the ILA to be transformed up to a more convenient value. To improve bandwidth (at the expense of antenna size), the wire running parallel to the ground may be replaced by a planar element, giving a PIFA. As shown in Figure 8.26**,** the sum of two adjacent sides of the PIFA should be approximately quarter of a wavelength long at the desired resonant frequency [19].

Evolution from a half-wave patch antenna is illustrated in Figure 8.27.

Here a probe-fed half-wave patch antenna is first short-circuited along a line where the electric field is zero (for the lowest frequency TM_{001} mode). This halves the size of the antenna without changing the resonant mode (though there is a loss of directivity, since radiation now occurs only from one end of the structure rather than two). Such a structure is often referred to as a *full short-circuit PIFA* (FS-PIFA), since the short circuit runs the full length of one face of the antenna (as indicated by the bold line in Figure 8.27). Replacing this short circuit with a simple shorting pin or tab reduces the resonant frequency further (by increasing the antenna inductance) and yields a PIFA.

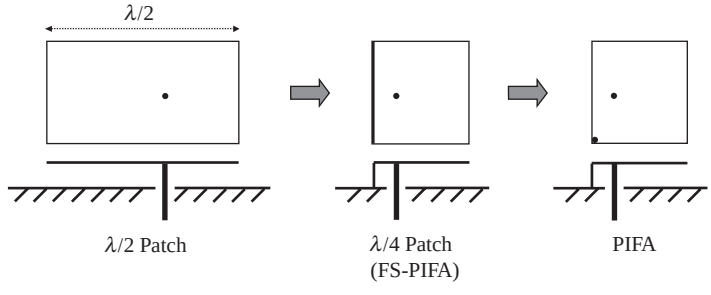


Figure 8.27 Evolution of a PIFA from a half-wave patch antenna. (Reproduced by permission of Delft University Press)

8.2.2.4 The Effect of the PCB

Many small antennas appear to exceed the fundamental limits of antenna size, bandwidth and efficiency outlined in Section 8.1.1. This is because they do not operate alone: there is a strong interaction with the equipment/PCB that tends to widen the bandwidth.

8.2.2.5 Specific Absorption Rate (SAR)

The SAR is an important parameter in the design of mobile phones because national and international regulations must be met and because consumers may buy a particular model of phone based on the quoted SAR.

SAR is regulated based on the known biological effects of thermal heating. The temperature increase of the body cannot be directly correlated with the incident radiation due to the complex thermoregulatory process within the body. Instead, the concept of power absorption per differential unit mass is used as follows:

$$SAR = \frac{dP}{dm} = \frac{dP}{\rho_D \, dV} \tag{8.50}$$

where,

dP	power absorbed within a differential volume
dm	mass within the volume
dV	differential volume
ρ_D	density

This can be expressed in terms of the electric field as follows:

$$SAR = \frac{\sigma \, |E|^2}{\rho_D} \tag{8.51}$$

where

σ	conductivity of the body (which will vary within different biological tissues)
E	the RMS electric field strength within the body

The SAR is a power relation specified at a particular point in space. To convert this to a temperature rise (and, hence, to a biological effect), it is necessary to average over both time and mass – or, equivalently, volume. Typical averaging masses of 10 g and 1 g are used in national and international regulations (unfortunately there is no unified world regulation). Averaging over 1 g gives significantly higher peak SAR values than averaging over a larger mass, since significant SAR variations occur on a small scale. Hence, the lower mass gives the most difficult specification and is used for most designs.

Figure 8.28 shows a PIFA and a generic monopole-like antenna, both mounted on a 100 $\times$ 40 $\times$ 1 mm PCB. These represent the two basic classes of internal antenna as described earlier in this section. The PIFA is a typical dual-band structure, while the monopole-like antenna is triangular – a structure that will have relatively low current density and, therefore, relatively low local field strength.

The antenna configurations are simulated next to a truncated flat representation of the phone user's head [20], as shown in Figure 8.29. Such a representation is often referred to as a

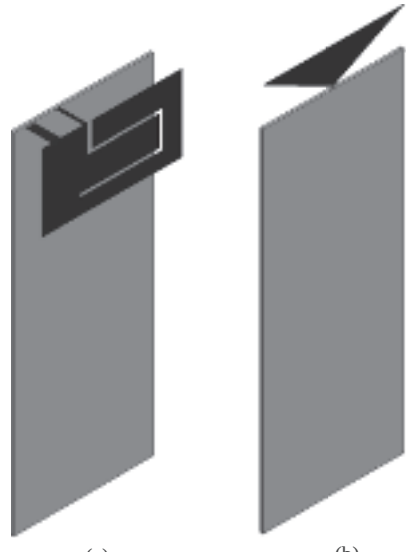


Figure 8.28 PIFA and generic monopole geometries (a) PIFA; (b) Monophone. (Reproduced by permission of Delft University Press)

Table 8.1 Relative permittivity and conductivity of phantom and skin layers

	Phantom		Skin	
Frequency (MHz)	Relative dielectric constant, ε_{pr}	Conductivity, σ_p (S/m)	Relative dielectric constant, ε_{sr}	Conductivity, σ_s (S/m)
900	41.5	0.9	4.2	0.0042
1800	40	1.4	4.2	0.0084

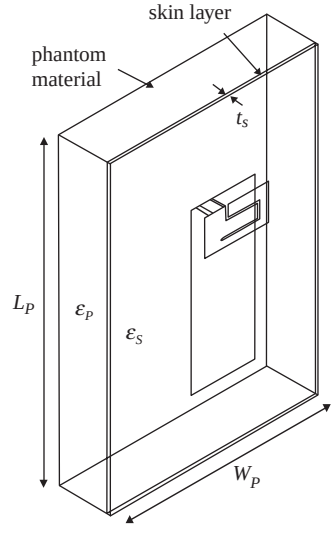


Figure 8.29 Flat phantom. (Reproduced by permission of Delft University Press)

Table 8.2 Intrinsic impedance as a function of frequency

Frequency (MHz)	Phantom impedance (Ω/square)	Skin impedance (Ω/square)
900	$54.35 + j12.06$	183.83
1800	$57.06 + j9.68$	

phantom. In both cases a separation of 5 mm is maintained between the PCB and the phantom surface. The material properties of the head and skin are taken from published databases.

To minimize reflections at the truncation surfaces of the phantom, these surfaces are defined as impedance boundaries, having the intrinsic impedances of the dielectrics used. Using Equation (3.19), the impedances are as given in Table 8.2

The simulated SAR of the PIFA antenna is shown in Figure 8.30, whereas the SAR of the monopole-like antenna is given in Figure 8.31. The SAR of the PIFA is clearly lower than that of the monopole-like antenna due to the shielding of the PCB and the increased distance of the antenna from the user. This is particularly the case at high frequencies when local fields at the antenna dominate.

This example shows that antenna conductors in the plane of the PCB give rise to high local SARs unless they can be moved further from the head. This effectively restricts the use of such

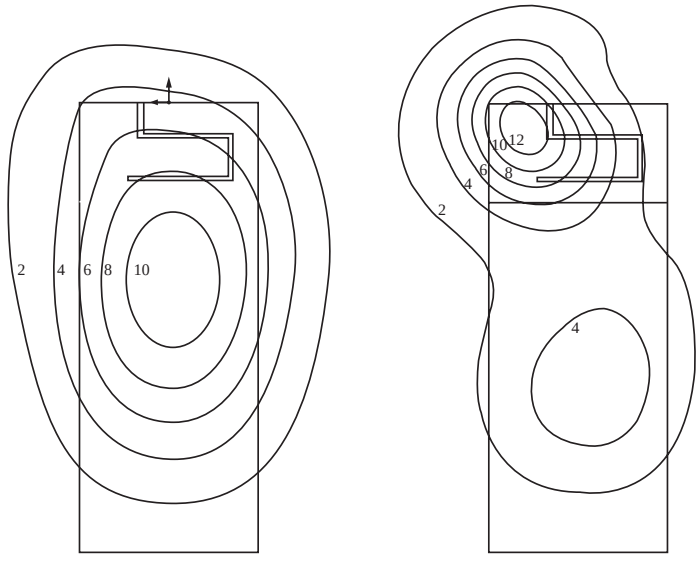


Figure 8.30 SAR in W/kg of a typical conventional PIFA (power normalized to 1 W) (a) 900 MHz; (b) 1800 MHz. (Reproduced by permission of Delft University Press)

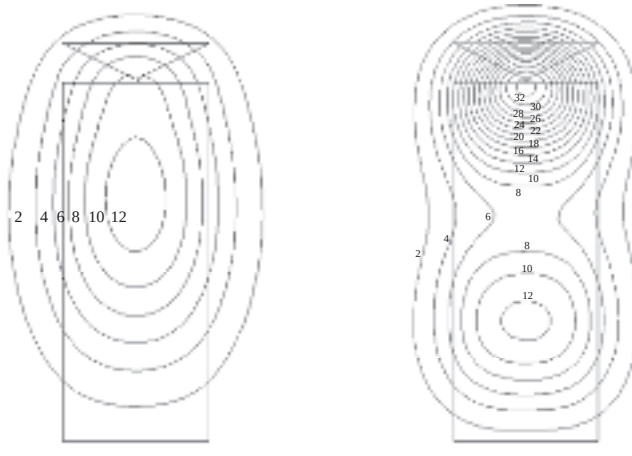


Figure 8.31 SAR in W/kg of a typical triangular monopole (power normalized to 1 W) (a) 900 MHz; (b) 1800 MHz. (Reproduced by permission of Delft University Press)

antennas to the bottom position of the phone. Only when the phone is long enough that there is sufficient clearance between the antenna and the head can monopole-like antennas be used with acceptable SAR. This is often the case for 'flip' phones.

8.2.2.6 Multipath and Mean Effective Gain

Multipath is the term used to describe the fact that radio signals often occur at a receiver via several different paths. This is illustrated for a typical mobile radio reception scenario in Figure 8.32. The scenario shown corresponds with that of a macro-cell, where an elevated

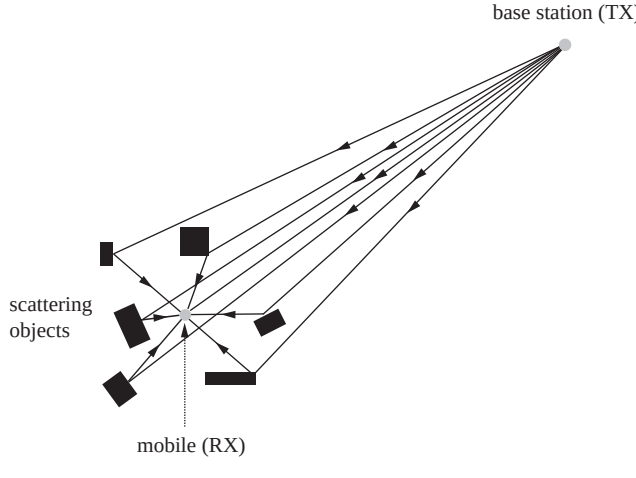


Figure 8.32 Typical macro-cellular multipath scattering scenario

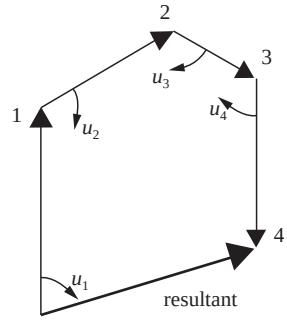


Figure 8.33 The addition of multipath vectors

base station is free of local scattering (reflections and re-radiation from nearby objects). The mobile, however, is below the height of local clutter and hence receives multipath contributions from many different angles. This is often referred to as a *ring of scatters* model, where typical radii of the scatterers from the mobile are of the order of a few hundred meters.

A summation of the multipath components occurs in the radio receiver, as shown in Figure 8.33. As the mobile moves, each multipath component vector rotates by a different amount, u_i, as indicated. Hence, constructive and destructive interference occurs – a phenomenon that is known as *signal fading*. This reflects the fact that the signal quality is very poor when the vector addition of the multipath components produces a *fade* (a low value due to destructive interference).

Fades tend to occur at intervals of approximately half a wavelength. Many readers will have experienced this. For example, FM radio has a wavelength of approximately three meters and is subject to fading. If a poor signal is experienced whilst listening in a car, reception can be improved by moving the car forward by a couple of meters.

The performance of an antenna in a multipath environment cannot be quantified simply in terms of its free space gain (radiation pattern) since the gain is only specified in one direction. Thus, we need a more accurate, environment-dependent term, the mean effective gain (MEG), given by

$$MEG = \frac{1}{(1 + X_P)} \iint_{\theta,\varphi} \left[G_\theta(\theta,\varphi)\, p_\theta(\theta,\varphi) + X_P G_\varphi(\theta,\varphi)\, p_\varphi(\theta,\varphi) \right] d\theta d\varphi \qquad (8.52)$$

where

G_θ and G_ϕ	are the antenna power gains in the θ and ϕ polarizations respectively.
X	is the cross-polar ratio P_ϕ/P_θ – the ratio of power received by θ and ϕ polarized isotropic antennas
p_θ and p_ϕ	are the angular density functions of the incoming θ and ϕ polarized plane waves, respectively

The gain includes all radiation-dependent parameters such as the antenna efficiency, the loss due to the proximity of the user and the radiation pattern (including the effect of the handset

Table 8.3 Typical model parameters for the global angle of arrival PDFs

Local environment	Cell type	m_θ	σ_θ	m_ϕ	σ_ϕ	X (dB)
Rural	Macro-cell	0°	15°	–	–	−15
Suburban	Macro-cell	15°	20°	20°	25°	−9
	Macro-cell	20°	30°	30°	50°	−7
Urban	Micro-cell	10°	25°	15°	30°	−7
	Pico-cell	5°	20°	0°	25°	−7
Indoor	BS outdoors	0°→ 5°	15°→20°	0°→ 5°	20°→25°	−2
	BS indoors (Pico-cell)	0°→ 5°	30°→40°	0°→ 5°	35°→45°	−2

and the user). The angular density functions and the cross-polar ratio define the propagation environment.

In any one direction, the gain of the antenna is weighted by the probability of receiving a multipath component from that direction. This weighting is performed for both polarizations. A further weighting is also applied based on the relative powers incident in both polarizations.

For high MEG, the antenna radiation pattern should be well matched to the angular probability distribution and cross-polar ratio of the incoming multipath. The azimuth (ϕ) distribution is assumed to be uniform, since multipath originates from all angles around the mobile. The elevation (θ, as measured from the ground upwards) distributions of some typical environments are given in Table 8.3. The distributions in elevation are considered here to follow a Gaussian distribution, given by

$$p_\theta(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\theta} \exp\left\{-\frac{(\theta - m_\theta)^2}{2\sigma_\theta^2}\right\} \tag{8.53}$$

$$p_\varphi(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\varphi} \exp\left\{-\frac{(\theta - m_\varphi)^2}{2\sigma_\varphi^2}\right\} \tag{8.54}$$

where

m_θ, m_ϕ, are the means of the angular probability density functions (PDFs) of the incoming θ and ϕ polarized plane waves, respectively.

σ_θ, σ_ϕ, are the standard deviations of the PDFs of the incoming θ and ϕ polarized plane waves, respectively.

Distributions other than Gaussian have also been proposed, though the exact form of the distribution is not normally very important to the final MEG. Polarization matching, for example, is much more important.

The shape of a typical θ-polarized urban macro-cell probability distribution is shown in Figure 8.34.

The antenna MEG is maximized when the antenna gain maximum coincides with the most likely direction of multipath reception.

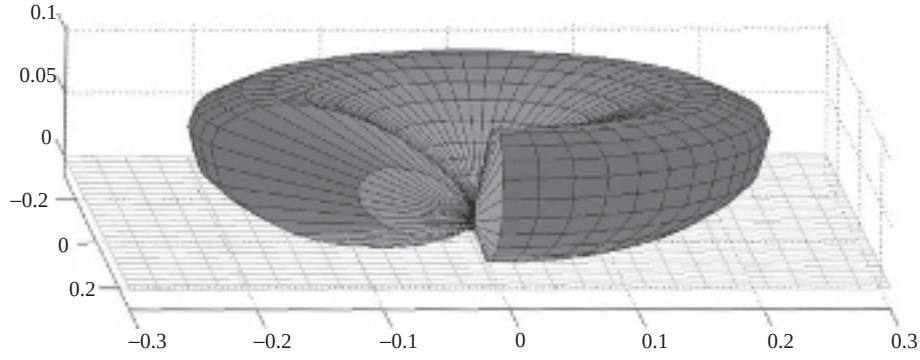


Figure 8.34 Angle of arrival PDF for an urban macro-cell copolarized with the base station

8.2.3 Antenna Diversity

Antenna diversity is a method in which more than one antenna is used to overcome the detrimental effects of multipath fading. Figure 8.35 shows the reception from two antennas subjected to the same multipath components. The signal fading from the two antennas occurs at different times/distances (time and distance can be considered to be interchangeable when the mobile moves at a constant velocity), since the antennas process or filter the multipath in different ways. By combining these two signals, the probability of a deep fade is much reduced.

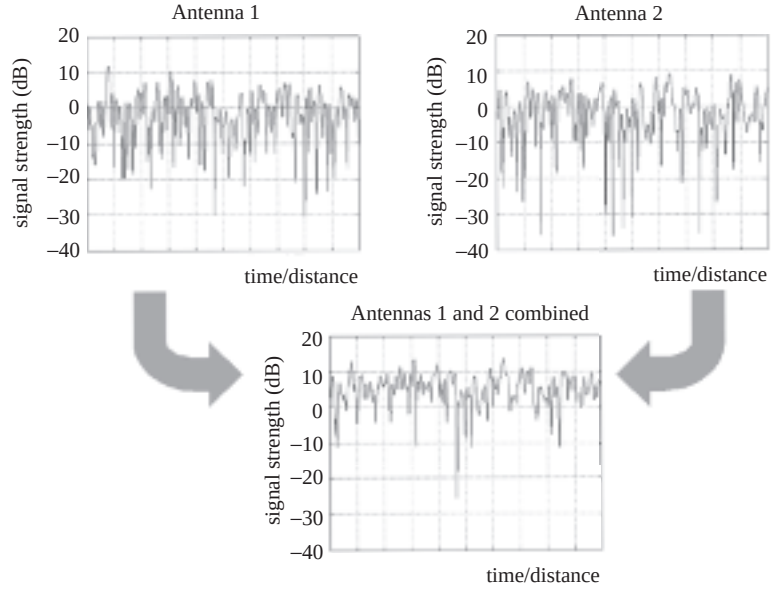


Figure 8.35 Combination of two uncorrelated signals

The fades shown in Figure 8.35 are typical of those experienced in narrowband cellular radio systems such as GSM and also those of noncellular systems such as FM radio. Occasionally, fades can cause the signal strength to drop by as much as 30 dB (both Antenna 1 and Antenna 2 have average signal levels of 0 dB in Figure 8.35). It can be seen that the depth of such fades and the probability that they will occur are reduced when the signals are combined.

To realize antenna diversity there are two requirements:

- signals must be decorrelated (to some degree) – i.e. two or more signals must be available that fade at different times/positions;
- a method of combining the signals must be found.

These two requirements are addressed in the sections that follow.

8.2.3.1 Decorrelation Methods

There are several ways of achieving decorrelated signals, i.e. signals that experience fades at different times or when the mobile is in a different position. These are often referred to as the types of antenna diversity described below.

1. **Polarization diversity:** antenna patterns are orthogonally polarized and hence preferentially receive only one of the two incoming polarization states. All multipath components will, to some degree, have both polarization states, the amplitudes and phases of which will be different. Hence, orthogonally polarized antennas will perform different vector summations of the multipath components and they will be decorrelated.
2. **Spatial diversity:** here multipath components will have different phases at the receive antennas due to their spacing. Unlike polarization diversity, the amplitudes of the multipath components will be the same at both antennas.
3. **Radiation pattern diversity:** here the magnitudes and phases of the antenna radiation patterns are arranged such that they do not overlap. Hence, the antenna directivity causes different multipath components to be added.

All three of the above mechanisms – polarization, spatial and radiation pattern – are usually present when antennas are closely spaced, such that decorrelated reception can be achieved at smaller spacings than often thought possible. Spacing is often limited by antenna isolation rather than antenna correlation.

The envelope correlation coefficient is given by

$$\rho_e = \frac{\left| \int_\Omega E_{1\theta} E_{2\theta}^* p_\theta \, d\Omega + X \int_\Omega E_{1\varphi} E_{2\varphi}^* p_\varphi \, d\Omega \right|^2}{\int_\Omega \left[|E_{1\theta}|^2 p_\theta + X \left| E_{1\varphi} \right|^2 p_\varphi \right] d\Omega \int_\Omega \left[|E_{2\theta}|^2 p_\theta + X \left| E_{2\varphi} \right|^2 p_\varphi \right] d\Omega} \tag{8.55}$$

where $E_{i\theta}$ and $E_{i\phi}$ indicate the complex radiation pattern of the ith antenna in the θ and ϕ polarizations respectively. It is important to note here that all radiation patterns have a

magnitude and phase in any given direction. Normally only the former is of interest, but the latter must also be considered when calculating correlation.

The envelope correlation coefficient is always between zero and unity. When zero, the signals are completely decorrelated (different) and when unity, the signals are completely correlated (identical).

8.2.3.2 Combining Methods

There are four common diversity combining methods, as follows (in order, with the worst-performing method first and the best last).

- **Switched combining (SWC):** here a switch is used to receive from one antenna at a time. When the signal received falls below a certain threshold, the second antenna is used. This does not guarantee better performance, but it does improve the likelihood. As suggested by the method, this is also referred to as threshold diversity.
- **Selection combining (SC):** the antenna with the strongest signal is used. This may be implemented with a single receiver or with two receivers. A single receiver is clearly desirable from a cost, size and power consumption perspective, however some time must be made available for the antenna signals to be compared. This is relatively straightforward in digital, packet-based systems. For example, DECT uses selection combining with the antenna selection performed in the preamble that occurs before each data packet is sent.
- **Equal gain combining (EGC):** here two receivers are required and the signals from the antennas are co-phased prior to summation.
- **Maximal ratio combining (MRC):** the signals from the antennas are weighted based on their SNR.

The differences between the methods are quite small for two antennas, as illustrated in Figure 8.36 for completely decorrelated antenna signals ($\rho_e = 0$). In this figure, the ordinate (y-axis) shows the probability that the signal-to-noise ratio is less than the level shown on the abscissa (x-axis) – i.e. the degree of certainty that a predefined signal-to-noise ratio will be achieved. For example, a CDF (cumulative distribution function) of 10^{-2} indicates a 99% probability that a particular signal level is exceeded.

When the CDF is low enough, the curves with diversity in Figure 8.36 become parallel lines and there is a constant relation between the diversity gains of the three combining methods. With two branches, EGC performs 0.88 dB better than SC, whereas MRC performs 1.5 dB better than SC. The diversity gain is substantial for all methods; however, the gain depends on the chosen certainty level. For example, for a CDF of 10^{-2}, an SNR of better than –20 dB is always achieved without diversity. With selection combining, an SNR of better than –10 dB is always achieved with the same level of certainty. Hence, the diversity gain is 10 dB. For lower levels of certainty (higher CDFs), the diversity gain will be lower than this.

For greater numbers of branches, the difference in diversity gain increases, particularly between MRC and SC. The variation in the diversity gain of MRC and SC with the number of branches is shown in Figure 8.37. EGC is not plotted, since an exact formula is not known for more than two branches.

The difference between SC and MRC with large numbers of branches is largely due to the fact that MRC with M branches requires M receivers, whereas SC is limited to one receiver.

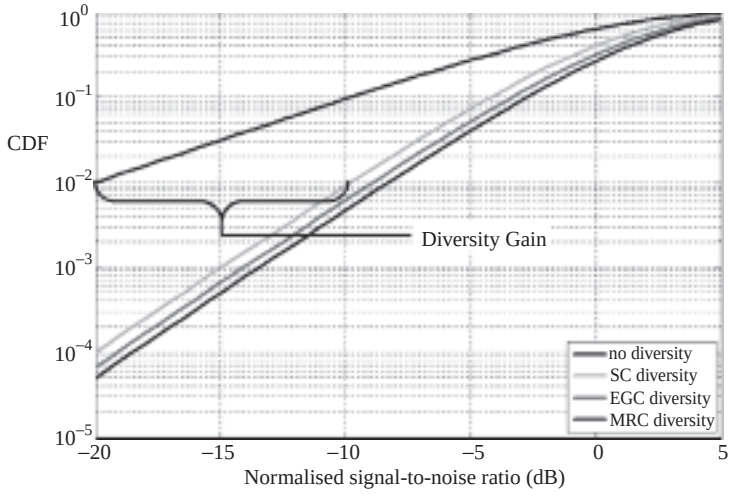


Figure 8.36 Diversity gains of SC, EGC and MRC with two equal power branches and zero correlation (Rayleigh signals). (Reproduced by permission of Delft University Press)

For all combining methods the mean power varies with the number of diversity branches. Assuming Rayleigh signal envelope statistics (valid for most multipath environments), zero correlation and equal branch powers, this variation is shown in Figure 8.38.

Clearly, both MRC and EGC give a linear improvement with the number of diversity branches. MRC gives the best performance but is only slightly better than EGC. SC is nonlinear

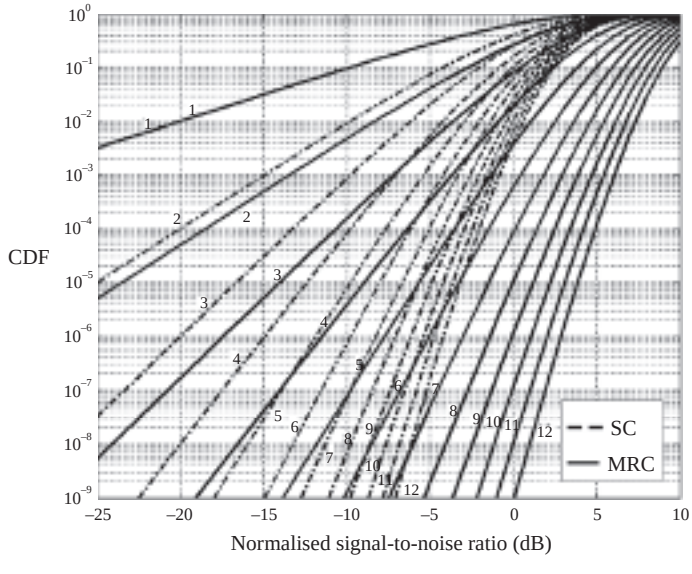


Figure 8.37 Diversity gains of MRC and SC with number of branches (1 to 12 branches shown from left to right). (Reproduced by permission of Delft University Press)

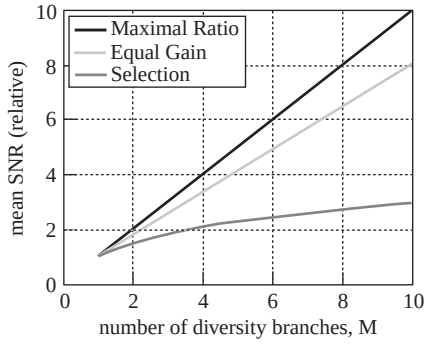


Figure 8.38 Mean SNR of different combining methods with the number of branches. (Reproduced by permission of Delft University Press)

and significantly worse than both MRC and EGC, particularly when the number of diversity branches is high.

8.2.3.3 The Effect of Branch Correlation

The effect of branch correlation for a two-branch MRC system with Rayleigh signals and equal branch powers is shown in Figure 8.39.

Figure 8.39 shows that there is some diversity gain even with completely correlated signals. This is simply a 3 dB gain that results from having two antennas rather than one (collecting twice as much of the incident energy). An appreciable additional diversity gain occurs at very high levels of correlation. Decreasing the correlation gives diminishing returns, such that it is often considered unnecessary to aim for a correlation coefficient of less than 0.5 to 0.7 (a correlation of 0.7 gives a diversity gain of ∼1.5 dB less than the optimum).

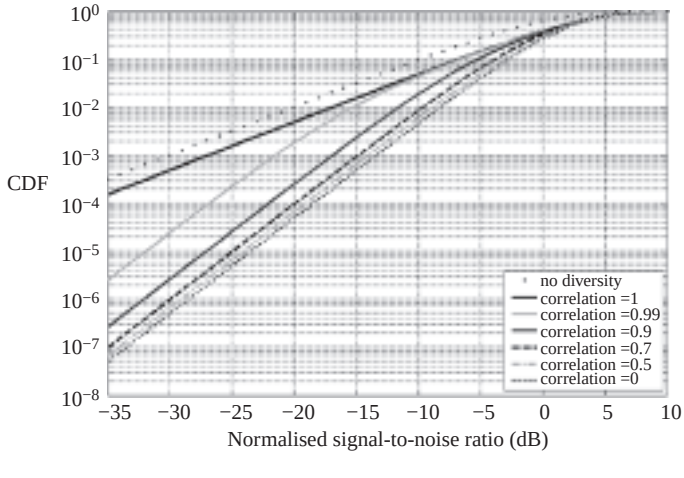


Figure 8.39 The effect of branch correlation. (Reproduced by permission of Delft University Press)

It can be seen that, as the CDF reduces, the curves for different correlation levels reduce at the same rate. Hence, the difference in diversity gain can be considered to be asymptotically constant.

8.2.3.4 The Effect of Unequal Branch Powers

The effect of unequal branch power is shown in Figure 8.40. For two branches, the diversity gain is nominally reduced by half the SNR difference between the branches.

$$Loss = \frac{SNR_1 - SNR_2}{2} \tag{8.56}$$

The power received by each antenna is given by the MEG.

8.2.3.5 Examples of Diversity Antennas

An example of two very simple diversity antennas within a DECT base station is shown in Figure 8.41. Here the base station is relatively large, such that low correlation can be guaranteed using space diversity: the antennas are separated by approximately 2/3 of a wavelength at the DECT center frequency of 1890 MHz. Simple quarter-wave wire monopole antennas are used to minimize costs.

An example of two diversity antennas within a DECT handset is shown in Figure 8.42. Here the antennas are arranged orthogonally in an attempt to realize polarization diversity. However, the handset PCB radiates significantly – in the same polarization – for both antennas, so the amount of polarization diversity that can be obtained is limited. Despite this, the differences in the radiation patterns of the two antennas are sufficient to ensure a usably low envelope

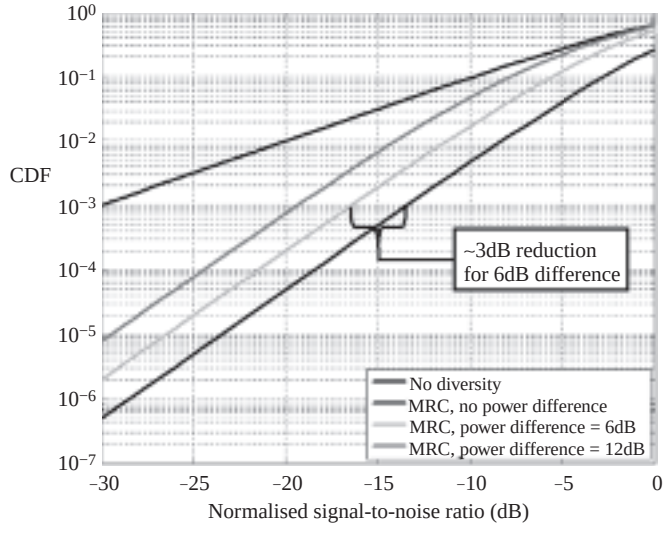


Figure 8.40 The effect of branch SNR differences. (Reproduced by permission of Delft University Press)

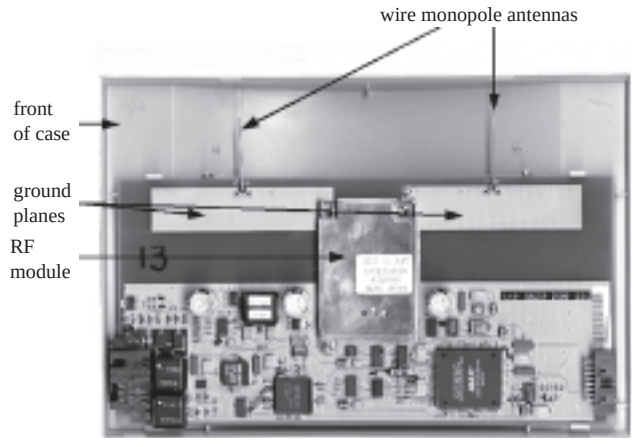


Figure 8.41 Diversity antennas in a DECT base station

correlation coefficient. The antennas used are inverted F antennas (IFAs) that are simply formed by tracks on the main PCB of the handset.

DECT uses SC with time made available in the preamble prior to packet transmission. IEEE 802.11b (Wi-Fi) also uses SC in a similar way to DECT. However, UMTS uses time diversity (i.e. there are alternatives to using antennas!).

8.2.3.6 MIMO Antennas

MIMO (multiple-in, multiple-out) antennas are emerging as a new technology to increase the capacity (data rate) of a communication system, whereas diversity antennas are used for improving the received signal quality and reliability. From an antenna point of view, MIMO antennas are similar to diversity antennas; they both require uncorrelated antennas and are most suitable for multipath environments where signals are random. But MIMO systems

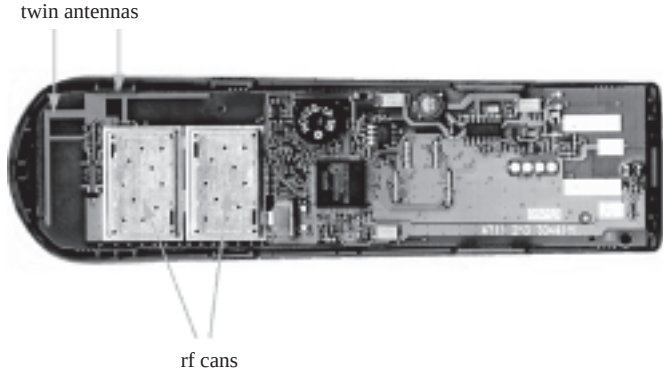


Figure 8.42 Diversity antennas in a DECT handset

need significantly more DSP (digital signal processing) and are much more complicated than diversity systems [21, 22].

8.2.4 User Interaction

8.2.4.1 Introduction

There is a strong interaction between antennas within portable devices and the user. This interaction causes the antenna resonant frequency to change – often to a lower frequency – and the antenna matching to vary. It also causes the antenna efficiency to reduce, since the body acts as a lossy dielectric. The only improvement is in the antenna bandwidth, as we would expect with reduced efficiency (refer to Section 8.1.1).

8.2.4.2 Body Materials

To consider the effect of user interaction on the performance of mobile phone antennas, we should first examine the material properties of the head. The variation of the relative permittivity of brain material with frequency is shown in Figure 8.43, whereas the loss tangent and skin depth are shown in Figure 8.44 and Figure 8.45 respectively.

The relative permittivity of brain matter varies slowly from approximately 40 to 50 between 10 and 1 GHz. Below 1 GHz it begins to rise increasingly quickly, reaching a value of approximately 80 at 100 MHz. The loss tangent also rises very quickly as the frequency is reduced below 1 GHz, reaching approximately 1.3 at 100 MHz. However, in contrast to the permittivity, the loss tangent is not monotonic and is a minimum at approximately 2.5 GHz.

Figure 8.45 shows that the skin depth rises sharply at frequencies below 1 GHz. The field penetrates to a depth of approximately 4.2 cm at the lower GSM frequencies at around 900 MHz, whereas at the higher GSM frequencies in the region of 1800 MHz, the penetration is approximately 2.7 cm.

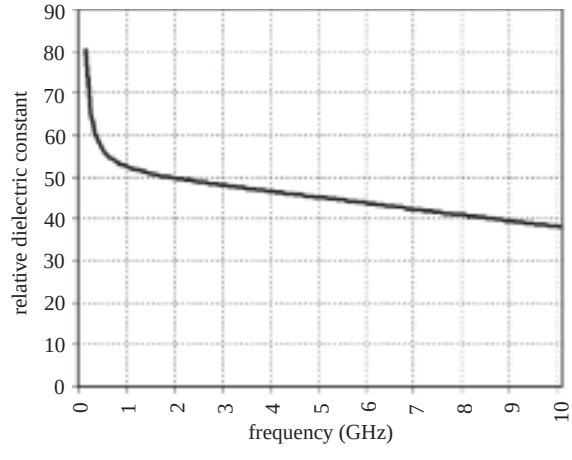


Figure 8.43 Permittivity of brain matter with frequency

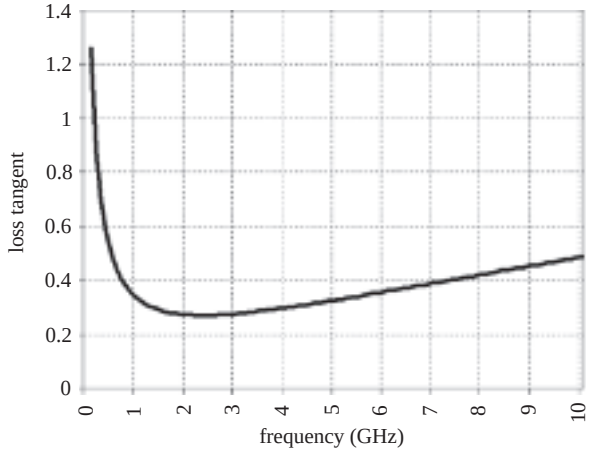


Figure 8.44 Loss tangent of brain matter with frequency

8.2.4.3 Typical Losses

The presence of the user affects the antenna performance in three main ways: the impedance match alters, the radiation efficiency is reduced and the MEG (introduced in Section 8.2.2) varies. Each of these mechanisms is addressed in the paragraphs that follow.

The radiation efficiency, η_R, reduces due to the additional losses in the body, particularly the head and hand. It is defined as the ratio of the power radiated (including all loss in the body) to the power accepted by the antenna.

The impedance change that occurs when a phone is held causes a change in the *mismatch loss*. This is expressed as unity minus the *mismatch efficiency*, $|\tau\ |^2$: the ratio of the power

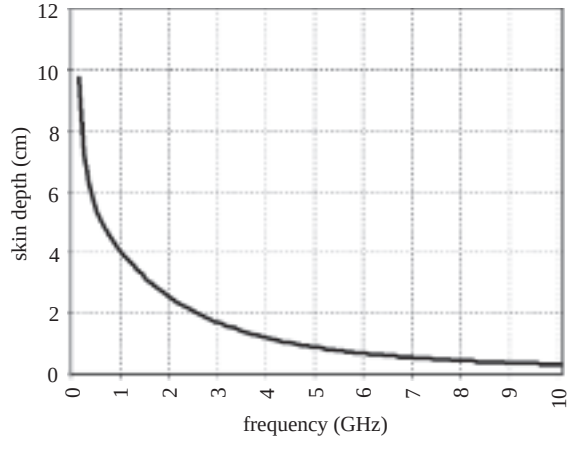


Figure 8.45 Skin depth of brain matter with frequency

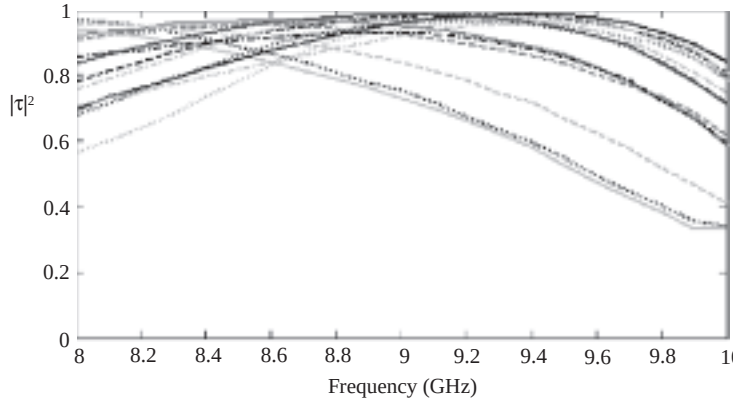


Figure 8.46 Typical mismatch efficiency with users in and around the GSM 900 MHz band. (Reproduced by permission of Delft University Press)

accepted by the antenna to the power supplied (the difference being the reflected power, due to impedance mismatch).

Figure 8.46 shows the typical mismatch efficiency of a mobile phone antenna in the 900 MHz band and in the presence of 15 different users, all of whom hold the phone in different ways. It can be seen that for nine users, the phone is not significantly detuned. For three users, the detuning is such that the resonant frequency is just out-of-band, although the band-edge mismatch efficiency remains high. For the remaining three users, the resonant frequency becomes significantly lower than is required, causing low mismatch efficiency at the upper band edge.

The total efficiency, η_T, is the ratio of the power supplied to the antenna to the integrated power on a sphere surrounding it (the radiated power). It is equal to the product of the radiation and mismatch efficiencies.

The total efficiency, mismatch efficiency and mismatch-corrected efficiency of four commercial mobile phone antennas (made in around 2000) in free space are shown in Table 8.4. The results are averaged over the GSM (880 MHz to 960 MHz) and DCS (1710 MHz to 1880 MHz) bands, using frequency steps of 10 MHz and 25 MHz respectively.

It can be seen that the best of these phones achieve average total efficiencies of the order of 60% at DCS and 70% at GSM in free space. The helical phone has a particularly poor return

Table 8.4 Average efficiencies of dual-band handsets in free space

	880–960 MHz			1710–1880 MHz		
Type	η_T (%)	$\lvert\tau\rvert^2$ (%)	η_C (%)	η_T (%)	$\lvert\tau\rvert^2$ (%)	η_C (%)
Helix	72.8	90.8	80.0	37.7	52.3	72.0
Flip	72.2	97.3	74.2	62.1	93.0	66.9
PIFA-1	70.4	95.2	73.6	64.2	90.9	71.1
PIFA-2	59.1	83.9	69.8	48.9	85.5	57.3

Table 8.5 Average efficiencies of dual-band handsets in the talk position

	880–960 MHz			1710–1880 MHz		
Type	η_T (%)	$\|\tau\|^2$ (%)	η_C (%)	η_T (%)	$\|\tau\|^2$ (%)	η_C (%)
Helix	5.3	71.7	6.9	4.9	50.8	9.8
Flip	5.7	90.2	6.1	5.8	91.5	6.3
PIFA-1	7.5	83.5	8.7	10.8	91.9	11.7
PIFA-2	3.4	55.2	5.7	10.7	85.0	12.7

loss at DCS and, therefore, a mismatch efficiency of approximately 50%. It is assumed that the match is improved (to some extent) using circuitry within the handset.

Typical efficiencies of dual-band phones in the talk position (held next to the head) are shown in Table 8.5. The average total efficiency is in the region of 3–8% in the 900 MHz band and 5–11% in the 1800 MHz band. The average radiation efficiencies are in the region of 6–9% and 6–13% in the 900 and 1800 MHz bands respectively. For both frequency bands, approximately half of the reduction in radiation efficiency can be attributed to the losses in the head, with the other half due to losses in the hand.

There is little difference between the average mismatch efficiencies in free space and in the presence of the user at 1800 MHz. However, for 900 MHz, the difference can be appreciable. In general, the resonant frequency reduces and the match at resonance deteriorates. However, the return loss bandwidth generally improves due to the losses induced in the body. In the 1800 MHz band, the net result is that the average stays approximately the same. In the 900 MHz band, the detuning of the resonant frequency generally outweighs any bandwidth enhancement caused by additional body loss (also shown in Figure 8.46).

The mean effective gain (MEG) is a function of the total efficiency, the radiation pattern and the multipath propagation statistics. It is the most representative calculable measure of performance in a real network. Additional loss can result if the radiation pattern changes in the presence of the body such that it becomes less well matched with the likely angles of arrival of multipath components than in free space.

The following summary gives the average MEG for 15 users at 920 MHz and 1800 MHz, calculated using the angle of arrival statistics given in Section 8.2.2. The mean and the standard deviation are given in Table 8.6 and Table 8.7 respectively.

The MEG is between 2.1% and 4.5%. This is significantly lower than the efficiency, partly because angles close to the horizon are most likely and the phones ‘waste power’ by having high relative gain outside the most likely angles. More importantly, however, the polarization of the phone radiation (reception) is not well matched to the polarization mix of the scattering environment.

The average MEG is –14.6 dB for both GSM and DCS. This represents a reduction of 1.8 dB for GSM and 3.5 dB for DCS when compared to the average efficiency. Clearly the radiation patterns at GSM must be better matched to the average propagation environment than at DCS.

Figure 8.47 shows the positioning of the users relative to the coordinate system chosen. Figure 8.48 shows the user-averaged radiation pattern of the flip phone, i.e. the gain at each angle is averaged for the 15 users. All users are right-handed, so only one average radiation

Table 8.6 Mean MEG (%) with 15 users (averages in parentheses are in dB)

	PIFA-1		PIFA-2		Helix		Flip	
	920 MHz	1800 MHz	920 MHz	1800 MHz	920 MHz	1800 MHz	920 MHz	1800 MHz
Rural	4.2	2.9	1.7	2.9	3.0	2.7	3.1	3.6
Suburban	4.2	3.9	1.9	3.8	3.0	2.2	3.3	3.6
Urban macro	4.4	4.2	2.0	4.1	3.1	2.1	3.4	3.6
Urban pico	4.4	3.8	2.0	3.8	3.4	2.5	3.6	3.9
Outdoor-to-indoor	4.8	3.8	2.7	3.8	4.4	2.5	4.6	3.9
Indoor	4.7	4.6	2.5	5.0	4.2	2.4	4.5	4.3
Average	4.5 (13.5)	4.0 (14.0)	2.1 (16.8)	4.1 (13.9)	3.5 (13.5)	2.4 (16.2)	3.7 (14.6)	3.9 (14.1)

pattern is required. Clearly, there is more power in the ϕ polarization than in the θ polarization. This is expected, due to the inclination of the phone when in a natural talking position. The blocking region of the head is clearly visible in the negative x, negative y direction. Negative x corresponds to the user's left-hand side, while negative y corresponds to the user's back. The left- and right-hand sides of the figure show the radiation pattern viewed from opposite directions in order to better show its 3D nature.

Figure 8.49 shows the user-averaged radiation pattern of PIFA-1. Again, the phone is predominantly ϕ polarized, but more so than for the flip phone. Both θ and ϕ polarizations have significant lobes in a near-zenith direction (along the positive z-axis). There is unlikely to be a propagation path supported in this direction, which also contributes to the inferior MEG performance of PIFA-1. Without this lobe, there is very little energy in the θ polarization. The propagation statistics used assume that this is the dominant polarization coming from the base station (via a scattering environment), which, again, results in low MEG. Assuming that the dominant polarization is θ directed implies a vertically polarized base-station antenna. It is worth noting that, as polarization diversity with slant 45° polarized antennas is becoming more common at the base station, this assumption may be questionable.

Table 8.7 Standard deviation of MEG (%) with 15 users

	PIFA-1		PIFA-2		Helix		Flip	
	920 MHz	1800 MHz	920 MHz	1800 MHz	920 MHz	1800 MHz	920 MHz	1800 MHz
Rural	1.9	1.5	1.5	2.3	2.7	3.0	2.1	3.3
Suburban	1.6	1.7	1.3	2.2	2.3	2.2	2.2	3.1
Urban macro	1.7	1.7	1.3	2.2	2.3	2.0	2.4	3.0
Urban pico	2.0	1.7	1.6	2.7	2.9	2.5	2.5	3.2
Outdoor–indoor	2.6	2.2	2.2	3.5	3.5	2.2	3.1	3.4
Indoor	2.5	2.1	2.1	3.4	3.3	2.1	3.0	3.4
Average	2.0	1.8	1.6	2.6	2.8	2.3	2.5	3.1

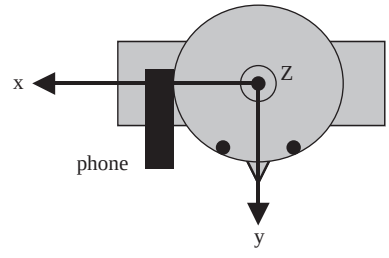


Figure 8.47 Orientation of users with respect to the chamber coordinate system. (Reproduced by permission of Delft University Press)

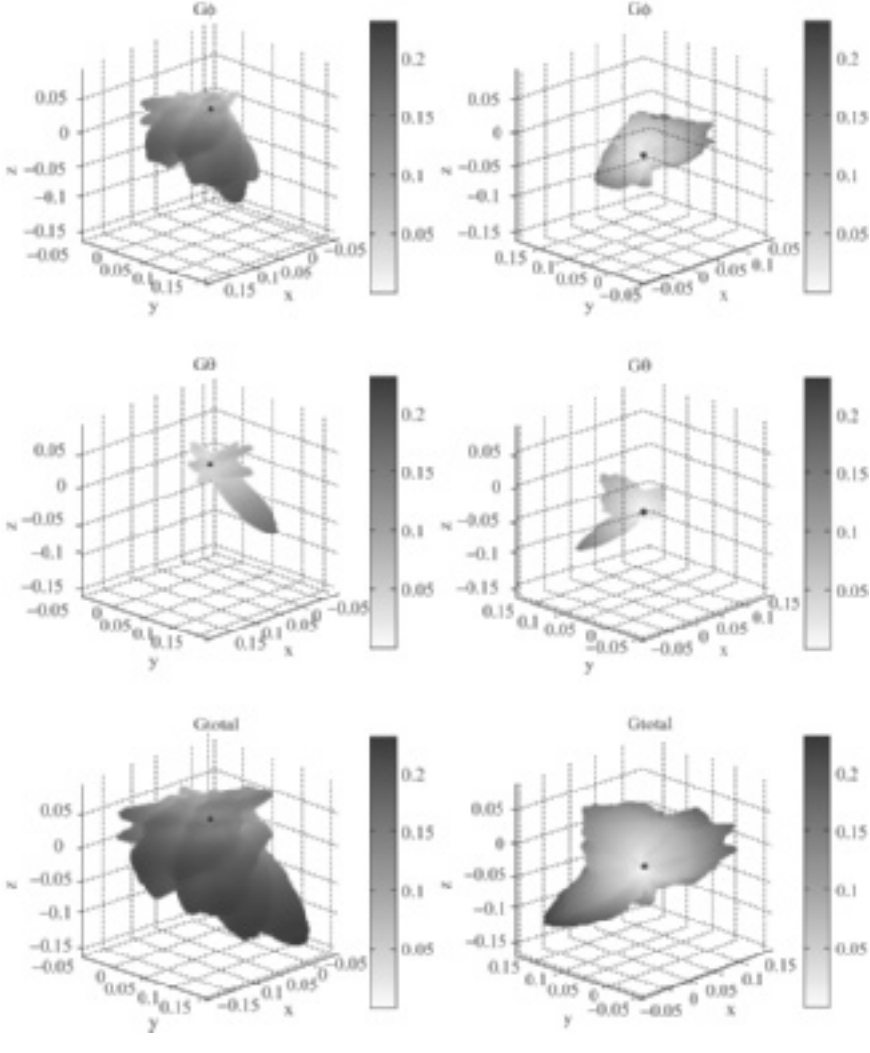


Figure 8.48 User-averaged radiation pattern of 'flip' at 1800 MHz. Radiation patterns on the left are viewed from an azimuth angle of 45°, while those on the right are viewed from an azimuth angle of −135°. The origin is marked by a small circle. (Reproduced by permission of Delft University Press)

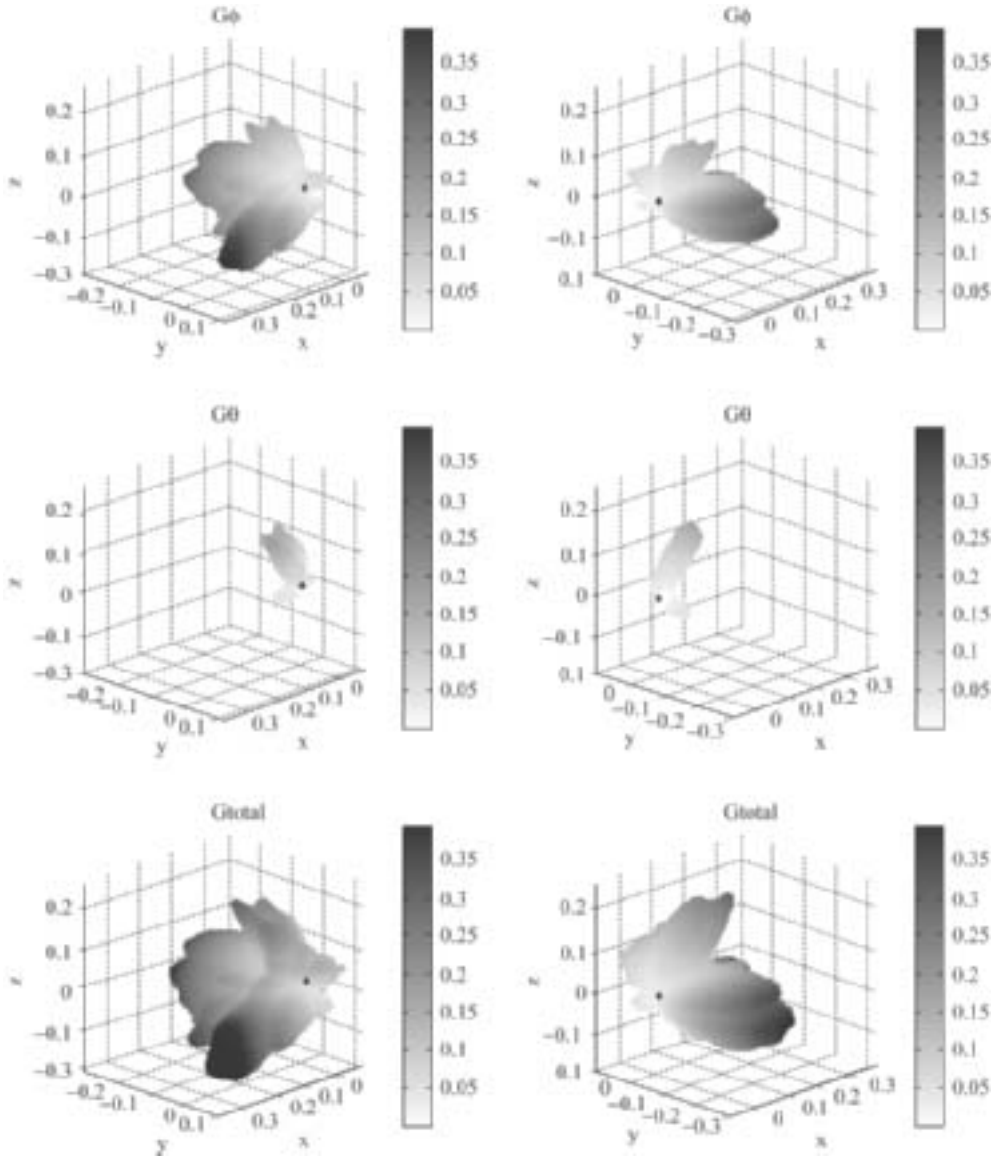


Figure 8.49 User-averaged radiation pattern of 'PIFA-1' at 1800 MHz. Radiation patterns on the left are viewed from an azimuth angle of 45°, while those on the right are viewed from an azimuth angle of −135°. The origin is marked by a small circle. (Reproduced by permission of Delft University Press)

Figure 8.50 shows the radiation pattern of PIFA-1 at the GSM center frequency, again averaged over all users.

At 920 MHz the polarization mix of all phones/antennas is more even than at 1800 MHz, which partly explains the superior relative MEG performance of the phones at GSM. The angular distribution of the patterns is also largely coincident with likely propagation paths. In

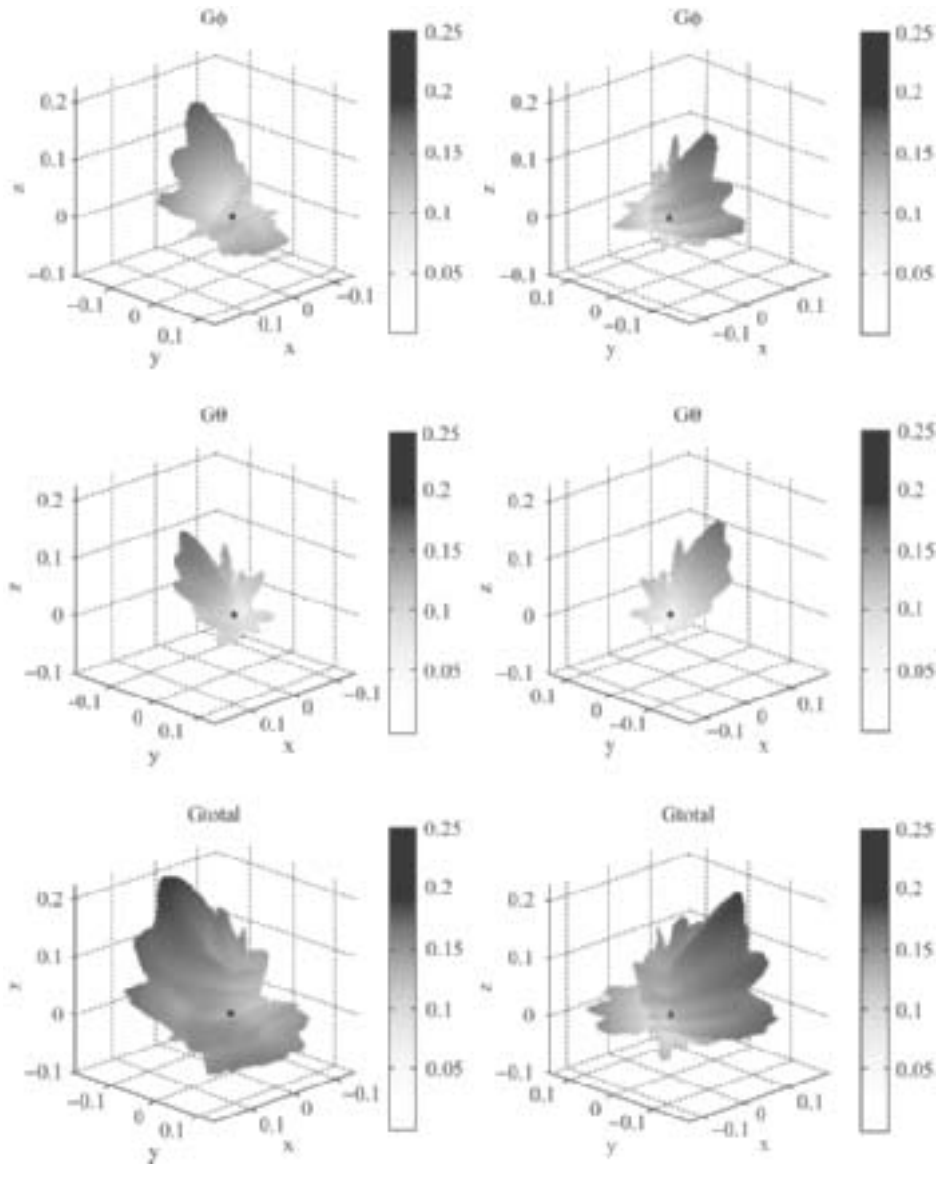


Figure 8.50 User-averaged radiation pattern of 'PIFA-1' at 920 MHz. Radiation patterns on the left are viewed from an azimuth angle of 45°, while those on the right are viewed from an azimuth angle of −135°. The origin is marked by a small circle. (Reproduced by permission of Delft University Press)

terms of MEG, all of the phones perform equivalently – within approximately 3 dB of each other at both frequencies.

8.3 Multiband and Ultra-Wideband Antennas

8.3.1 Introduction

Many modern communication systems have to work well over multiple frequency bands. For example, a well-designed cellular mobile phone is normally expected to operate in at least the two GSM bands (900 MHz and 1800 MHz) and the PCS band (1900 MHz). The recent release of the UWB band (from 3.1 to 10.6 GHz) for wireless low-power applications has generated a lot of interest in developing UWB systems [23, 24]. In this section we are going to deal with these two types of antennas.

8.3.2 Multiband Antennas

8.3.2.1 Techniques

Several basic principles can be employed to realize multiband antennas, as follows:

- use higher order resonances;
- use resonant traps;
- combine resonant structures;
- use parasitic resonators.

Each will be addressed in the sections that follow.

Higher Order Resonances

The use of higher order resonances is illustrated in Figure 8.51, which shows the resonant modes of monopole antennas as their length is increased in $\lambda/4$ increments.

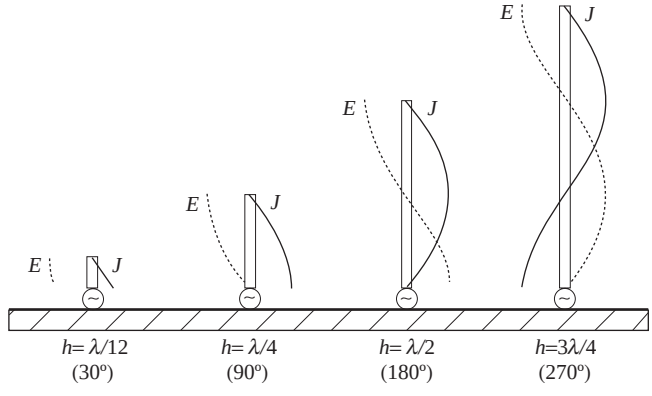


Figure 8.51 Resonances of monopole antennas of increasing length. *E* and *J* indicate the electric field and current density magnitudes respectively

Monopoles are often used with a length of λ/4 when the ***E*** field at the feed is a minimum and the current is a maximum. A similar condition exists at the feed when the antenna is 3λ/4. Alternatively, we can say that, for a fixed antenna height, the feed conditions will be similar when the frequency is equivalent to heights of λ/4 and 3λ/4: the antenna will be resonant at a first frequency and then a second frequency three times the first. Other natural resonances will also exist at higher frequencies.

Higher order resonances are present in many types of resonant antennas such as patches, dipoles, monopoles, slots and dielectric resonators and are often used (and manipulated) to give multiband operation.

Resonant Traps

Figure 8.52 shows a monopole antenna with a parallel resonant circuit – or *trap* – located approximately halfway along its length. The trap enables the antenna to work at two frequencies: a low frequency, f_1, and a higher frequency, f_2. At f_2 the trap is tuned to be close to anti-resonance (a high impedance) and current flow is restricted to the lower part of the antenna; resonance is achieved when the lower part of the antenna is close to λ/4 long. In practice, this resonance is achieved when this part is slightly less than λ/4 and the trap is anti-resonant at a slightly lower frequency than f_2. At f_1 the trap is well below resonance and is inductive, and the antenna is resonant when the full length is close to λ/4. Again, due to the shortening effect of the inductor, in practice resonance is achieved at a length that is slightly less than λ/4.

Radio amateurs use traps in dipole antennas located horizontally above ground, for example to cover the 80 and 40 meter bands (3.5 to 4.0 MHz and 7.0 to 7.3 MHz respectively). They are also often used in automotive antennas.

Combined Resonant Structures

Two or more resonant structures can be closely located or even co-located with a single feed point in order to achieve multiband operation. This is illustrated in Figure 8.53(a), which shows two closely spaced, adjacently located monopoles with a common feed point. The larger element is operational at the lower frequency, f_1, whereas the smaller element is operational at the higher frequency, f_2.

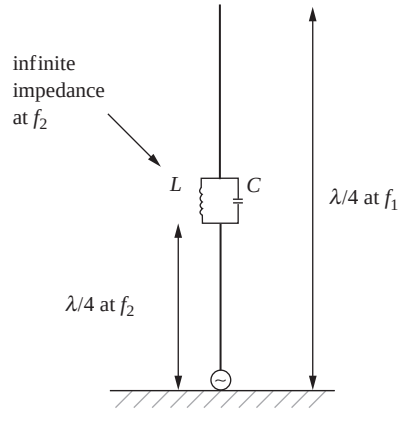


Figure 8.52 A monoopole antenna with a resonant trap

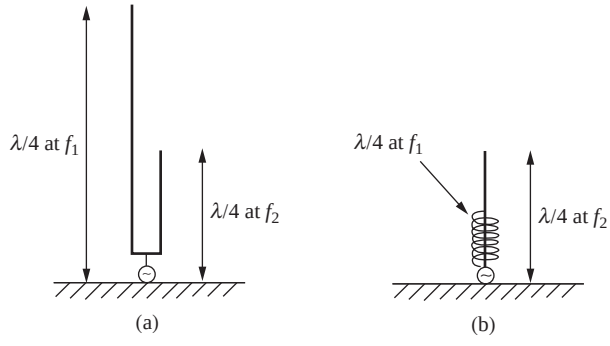


Figure 8.53 Combined resonant structures

Figure 8.53(b) works using the same principle; however, the low-frequency element is realized using a helical element to reduce space (the high-frequency element is within the helix for some of its length).

Parasitic Resonators

In contrast to the previous section, we may add a second element for operation at a second frequency, but without feeding it directly. This is illustrated in Figure 8.54, which shows two monopoles, one of which is fed, whereas the other is parasitically coupled via the near field of the fed antenna. Here the parasitic antenna is shown with a load, which can be used for tuning. Usually, the load would be reactive in order to maintain high antenna efficiency.

8.3.2.2 Examples

The first example uses the principle of higher order resonances. Figure 8.55 shows a simple normal mode helix antenna mounted on a PCB of typical mobile phone dimensions. The antenna is first resonant at a frequency, f_1, of approximately 0.84 GHz, with a resistance of

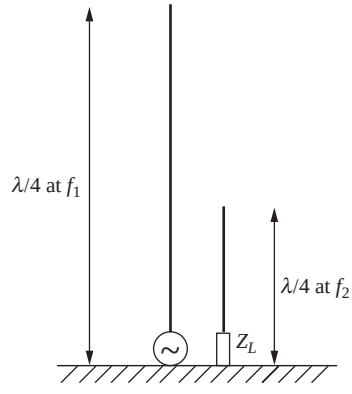


Figure 8.54 Parasitically coupled monopole antennas

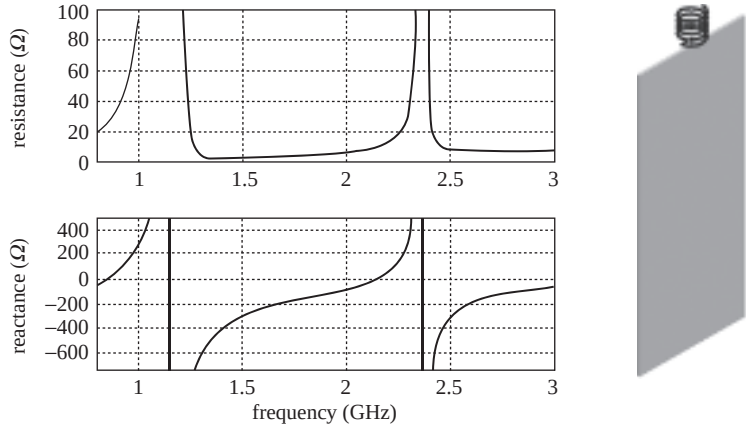


Figure 8.55 A small helical antenna mounted on a mobile phone PCB

approximately 27 ohms. At 1.16 GHz the antenna is anti-resonant (when the resistance is very high and the reactance makes the transition from being positive to negative). The antenna then becomes resonant again at a frequency, f_2, of approximately 2.15 GHz – when the resistance is 10 ohms – before quickly becoming anti-resonant again at approximately 2.35 GHz. The ratio between the first and second resonant frequencies is given by

$$\frac{f_1}{f_2} = 2.6 \tag{8.57}$$

This is close to the value of three that we would expect for resonances at electrical lengths of $\lambda/4$ and $3\lambda/4$ (as illustrated in Figure 8.51).

Since the antenna shown in Figure 8.55 is electrically small, the reactance can be seen to vary quickly at the first resonance and even more quickly at the second resonance. This is indicative of narrow bandwidth.

For dual-band mobile phones, the antenna resistance is normally required to be close to 50 ohms and the frequency ratio is close to two. A simple method of altering the frequency ratio is to have different helix pitches in the upper and lower portions of the antenna. This is illustrated in Figure 8.56 for a commercial dual-band helix. The low-frequency resonance is controlled by the overall length of the structure, whilst the high-frequency resonance is controlled by the pitch of the upper part of the antenna.

Alternatively, Figure 8.57 illustrates a dual-band helical antenna in which two windings of different pitches and lengths are integrated into a single volume with a common feed – an example of combined resonant structures.

8.3.3 Wideband Antennas

There is a wide variety of wideband antennas, many of which are also electrically large. Here we focus on wideband antennas that are also electrically relatively small, since this is often an industrial requirement.

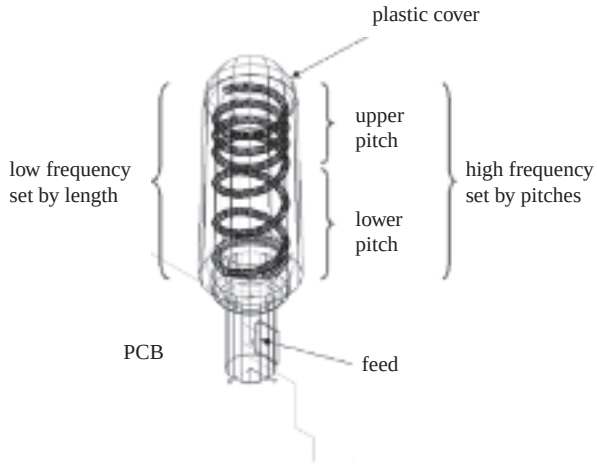


Figure 8.56 A dual-band helical antenna with two pitches for control of the ratio of the resonant frequencies

The basic approach to making an electrically small antenna wideband is to make it fat. We saw previously in Section 8.1.1 that the bandwidth of an antenna is related to the size of the sphere that just encloses it. By making the antenna fat, more of the sphere is occupied and the antenna bandwidth can be maximized. Some typical fat monopoles are illustrated in Figure 8.58. These and other similar shapes have formed the basis for UWB antenna designs.

UWB antennas are normally planar. However, as an example of what happens to the feed impedance when an antenna is 'fattened', we will concentrate on a three-dimensional conical monopole, which subtends an angle α at the feed. The resistance of such an antenna with height and flare angle is shown in Figure 8.59. When the monopole is thin – i.e. for $\alpha = 10^{\circ}$ – there is a strong resistance peak at approximately 124 degrees. For a very thin monopole, this peak would occur at 180 degrees, when the antenna is half-wave. However, as

Figure 8.57 A dual-band helical antenna utilizing two different length windings with a common feed point

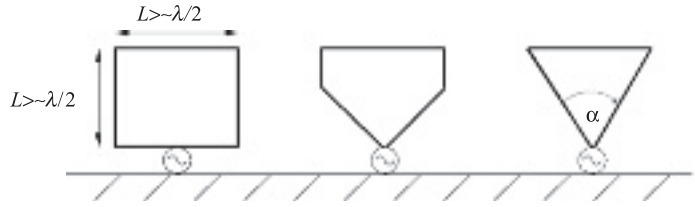


Figure 8.58 Some 'fat monopole' antennas

the flare angle increases, the half-wave peak occurs at progressively lower electrical heights and with a decreasing resistance value. When the flare angle reaches 90 degrees, the resistance is close to 50 ohms over a wide range of electrical heights. If the electrical height is greater than approximately 50 degrees, the resistance is close to 50 ohms over a wide frequency range.

The reactance of the conical monopole with height and flare angle is shown in Figure 8.60. As α is increased, the reactance variation reduces and is close to zero for heights greater than 90 degrees. For heights greater than 50 degrees, the reactance is within, or capable of being matched to within, acceptable limits.

We can conclude that monopole antennas with wide flare angles can exhibit near-constant resistance with electrical heights greater than approximately 50 degrees. In other words, small, fat antennas can exhibit wideband characteristics. Some UWB antennas have been developed based on this idea [24]. For example, two UWB antennas are shown in Figure 2.34 and the S_{11} of Antenna 1 is given in Figure 8.61, which has a return loss greater than 10 dB over the UWB band.

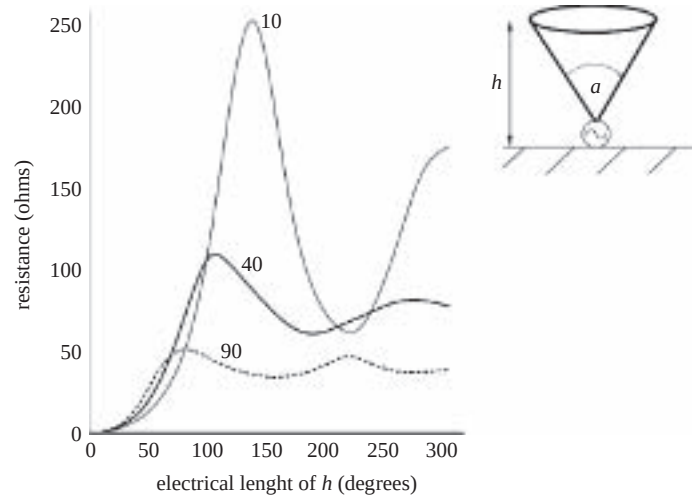


Figure 8.59 Resistance of a conical monopole with electrical length and flare angle (in degrees)

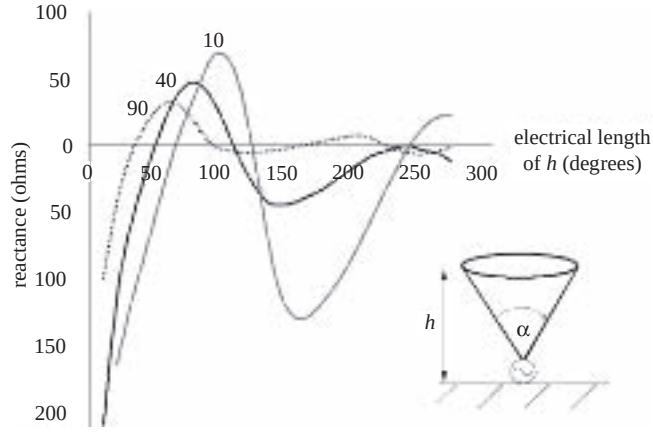


Figure 8.60 Reactance of a conical monopole with electrical length and flare angle (in degrees)

8.4 RFID Antennas

8.4.1 Introduction

Radio frequency identification (RFID) systems are short-range – normally digital – wireless systems that, as the name suggests, are used primarily for identification purposes [25]. Examples of applications include: animal tagging, asset tracking, electronic passports, smartcards and shop security. A simplified diagram of an RFID system is shown in Figure 8.62.

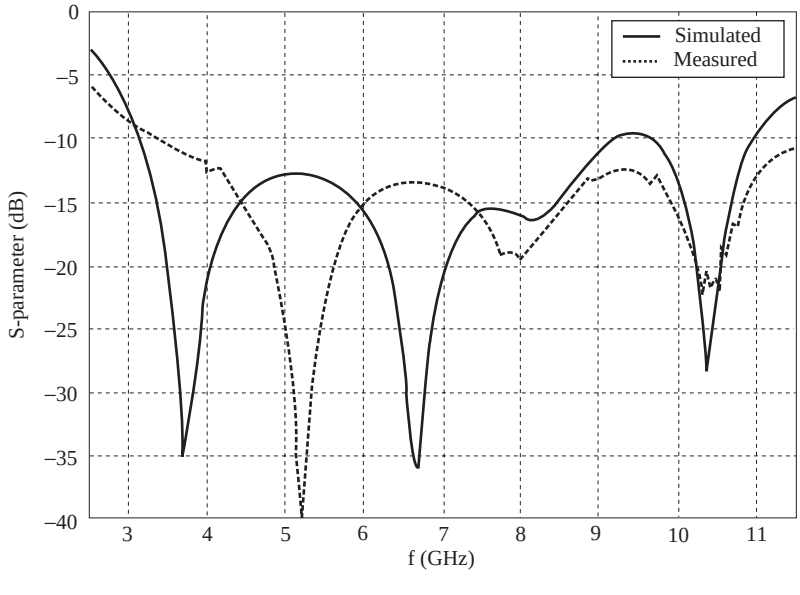


Figure 8.61 S_{11} in dB of a typical UWB antenna

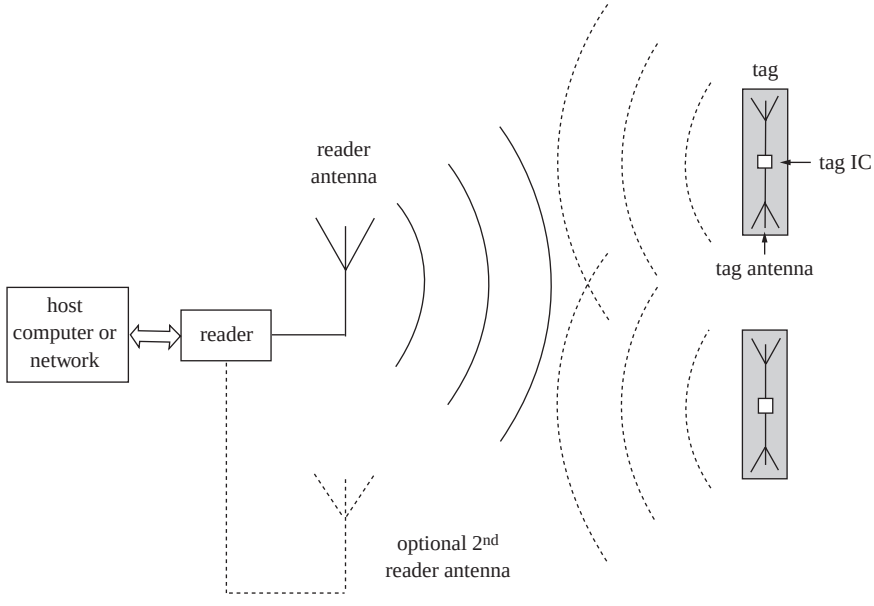


Figure 8.62 A simplified RFID system

At one end of the radio link is a small device with limited intelligence: the tag or transponder. Tags are designed for manufacture in very large volumes at very low cost. They can be likened to bar-code labels: they have an electronic product code (EPC), comparable to the universal product code (UPC) format used by bar codes, but they are also programmable and have the capability to store user-specific information (i.e. they contain some user memory). Tags contain an integrated circuit (IC) that is connected to a thin planar sheet that serves as a label and as a substrate for the antenna. The tag antenna is normally balanced (to suit the differential inputs of the tag IC), electrically small and linearly polarized.

At the other end of the link is a more expensive and sophisticated device: the reader or interrogator. These commonly used names reflect the fact that the primary jobs of the reader are to prompt communication with the tags (often with several tags simultaneously) and then to receive data from them. However, contrary to the name, the reader often also transfers data, or writes, to the tag. Before prompting the tag to respond, the reader also often supplies it with power; this helps to reduce the cost of the tags, since a battery is expensive to supply, connect and maintain.

There are several classes of tag, dependent on how they generate enough power to communicate, as follows:

- passive;
- semi-active;
- active.

Passive tags are very low in cost but tend to be limited in terms of range and data rate. They can only respond once powered by the reader, which may take some time. Semi-active

tags have a limited power supply, which allows faster operation. Active tags have enough on-board battery power to allow 'broadcast' of their presence to readers and high data rate communications.

The reader may be either *monostatic*, where transmission and reception uses the same antenna, or *bistatic*, where the transmission and reception antennas (and associated circuitry) are separate (as shown in Figure 8.62 with the second antenna used for receive only). The first and second antennas may also both be used for transmit and receive, for example where the reader antennas are installed on either side of a 'gate' through which tags pass (commonly used in shop security systems).

The reader antenna is generally electrically larger than the tag antenna and may be either balanced or unbalanced. Many different types of antenna can be utilized, with radiation patterns and polarization tailored for the particular application. Reader antennas are often circularly polarized when the orientation of the tag is unknown. While this reduces the maximum possible power transfer between the reader and the tag (there is a 3 dB coupling loss between circularly and linearly polarized antennas), it eliminates any likelihood that polarizations could be completely mismatched. This is important in tagging applications, where reliability is required to be very high.

Figure 8.63 shows a typical sequence of events for a passive tag to be identified by a reader. First, the tag transmits a continuous wave (CW) signal that is received by nearby tags. This signal is rectified to provide enough power to operate the tag. The reader then sends a modulated request for active tags (i.e. tags close enough to the reader to be powered up) to respond. This is followed by a response from the appropriate tag. Tag modulation is achieved by switching the load impedance seen at the terminals of the tag antenna. This causes the tag antenna to alter the field seen at the reader antenna. Although the change is very small, it can be detected and demodulated by the reader. In this way, the tag uses as little power as possible in order to maximize the read range: it does not generate a signal, but simply provides enough power to activate a switch.

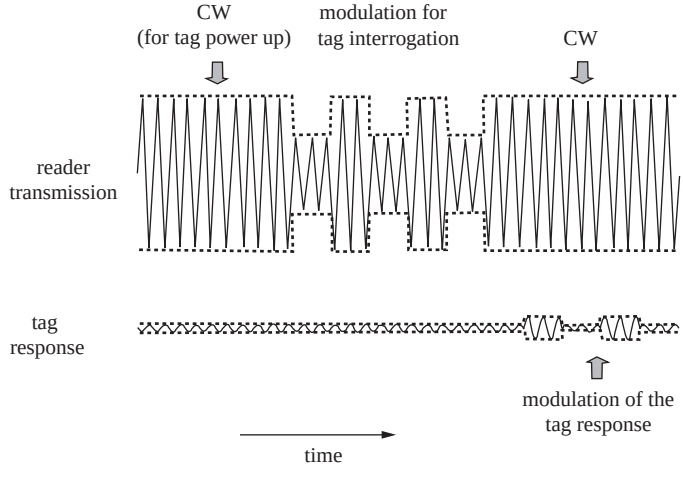


Figure 8.63 A typical sequence of events for a tag to be powered, interrogated and read

RFID systems operate at a number of frequency bands that can be grouped as follows:

- low frequency (LF);
- high frequency (HF);
- ultra high frequency (UHF).

There are large differences in the frequencies used in the LF, MF and UHF bands. Because of this, there are also significant differences in the way in which the systems operate and in the applications for which they are appropriate.

LF and HF systems use near-field coupling between the tag and the reader antennas: most often the coupling is inductive, but it may also be capacitive. When inductive coupling is used, both the reader and the tag typically use multi-turn coils as antennas. In fact, referring to these coils as antennas is a little misleading, since the coils of the reader and tag act more like transformers: the coil of the reader can be considered to be the primary winding and the coil of the tag acts as the secondary. The coupling between the coils (and therefore the transformation ratio) is dependent on their separation and always occurs within the near-field region (where the magnetic field falls with the inverse of distance to the power of three). Near-field systems are discussed in detail in Section 8.4.2. They are used particularly in applications where the tag may be attached to a lossy dielectric such as a fluid-filled bottle, an animal or a solid object. This is because the tag antenna electrical characteristics tend to change less than for far-field (UHF) antennas.

UHF systems generally use far-field coupling, when the antennas of the reader and tag operate in a more conventional way: the transfer of energy is determined by the Friis equation (see Section 3.5.1). UHF far-field systems typically have greater range than LF and HF systems. As such, they are often used in systems where higher than normal levels of performance are required. They are described in greater detail in Section 8.4.3.

8.4.2 Near-Field Systems

For near-field systems, loop/coil antennas are normally used for the reader and the tag. Consider the simple reader loop shown in Figure 8.64, where a is the reader loop radius, b is the tag loop radius and d is the separation between the tag and reader loops. Close to the reader the magnetic field produced is greatest along the axis of the loop: the z-axis as drawn. Hence, products are often arranged so that the reader and tag loops have the same axis, as shown.

With such an alignment, the mutual inductance, M, between the loops is given by [26]

$$M = \mu\sqrt{ab}\left\{\left(\frac{2}{D} - D\right)K(D) - \frac{2}{D}E(D)\right\} \tag{8.58}$$

where

$$D^2 = \frac{4ab}{d^2 + (a+b)^2} \tag{8.59}$$

$$E(D) = \int_0^{\pi/2} \sqrt{1 - D^2\sin^2\theta}\, d\theta \tag{8.60}$$

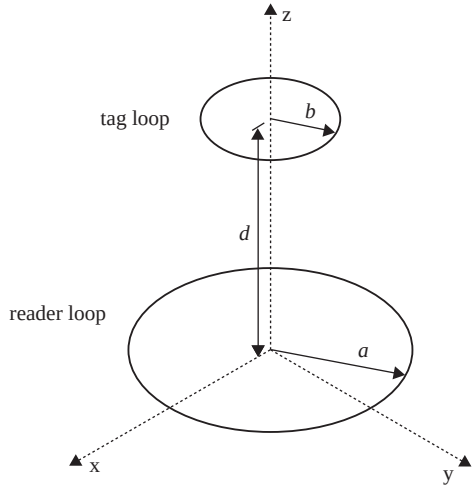


Figure 8.64 Simplified reader and tag loop antennas

$$K(D) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - D^2 \sin^2 \theta}} \tag{8.61}$$

$E(D)$ and $K(D)$ are complete elliptic integrals of the first and second kind, respectively. These formulas indicate that there is an optimum relationship between the sizes of the tag and loop antennas that should be considered at a system design stage.

A typical practical near-field system is shown in Figure 8.65. Here the tag antenna size and geometry is normally limited by the application: for example, credit card dimensions are often used for smart cards. Multi-turn loops are also often employed, when the mutual inductance increases linearly with the number of turns.

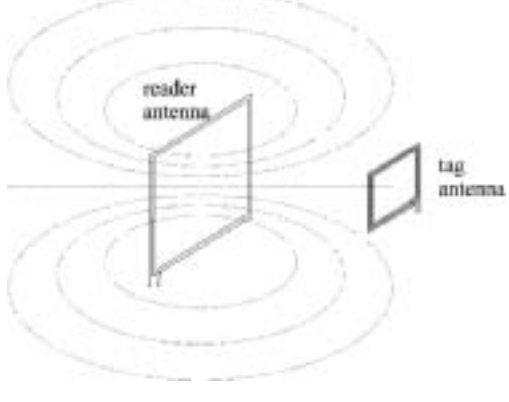


Figure 8.65 A near-field RFID system

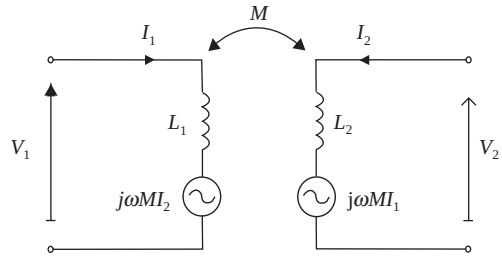


Figure 8.66 Mutual coupling between reader and tag coils

The transfer of energy between the reader and the tag is via the mutual coupling between the two loops. The equivalent circuit of two mutually coupled inductors is shown in Figure 8.66.

The circuit on the left of Figure 8.66 represents the reader, whereas the circuit on the right represents the tag. The transfer of energy that occurs between the two coils is modeled by two current-dependent voltage sources connected in series with the coil inductances. These sources depend on the level of mutual coupling, which varies with the distance between the tag and reader. The mutual coupling is typically low, such that the reader current can be considered to be approximately constant. This allows the tag to be analyzed independently of the reader.

In addition to its inductance, the tag loop will have some associated loss resistance and some inter-turn capacitance, which is most simply approximated by a shunt capacitor. The tag IC will have a load resistance and a capacitance that is designed to be appropriate for certain loop dimensions and frequency ranges. The IC is normally modeled as either a series or parallel equivalent circuit. The equivalent circuits of a practical tag inductor and a tag IC are shown in Figure 8.67.

In the example given in Figure 8.66, the IC is modeled as a parallel equivalent circuit, whereas the inductance and resistance of the tag loop are modeled as a series equivalent circuit. To make

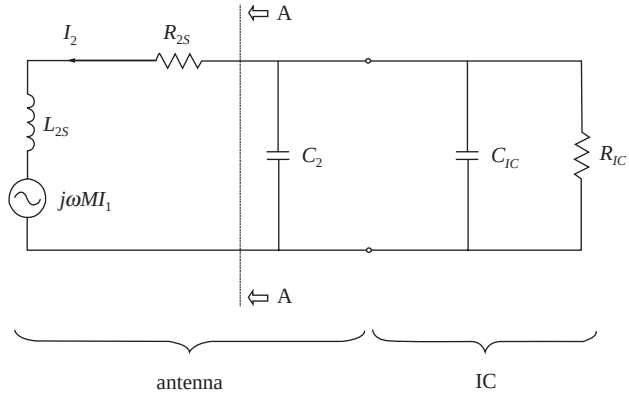


Figure 8.67 Equivalent circuit of a tag antenna and IC

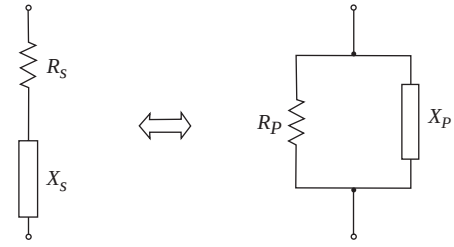


Figure 8.68 Equivalent circuit of a tag antenna and IC

the system easier to analyze, it is necessary to obtain a completely series or parallel equivalent circuit. This can be done easily by using Norton or Thevenin equivalents as appropriate. In the example that follows we will convert components to the left of A–A to a Norton (parallel) equivalent circuit. To do this, we short-circuit any independent voltage sources (the source in Figure 8.66 can be considered independent if the mutual coupling is low and I_1 can, therefore, be considered to be constant) and determine the impedance looking into the circuit at A–A. We also short-circuit the terminals at A–A, where the current becomes that of the equivalent Norton current source. The Norton impedance is given by $R_2 + j\omega L_2$. To convert this to parallel connected components we use the equivalent circuit shown in Figure 8.68.

Series-connected resistance and reactance, R_S and X_S respectively, can be represented at a single frequency by a shunt-connected resistance and reactance, R_P and X_P respectively. The relationships between components are

$$R_P = \left(\mathbf{1} + Q^2\right) R_S \tag{8.62}$$

and

$$X_P = \frac{\left(1 + Q^2\right)}{Q^2} X_S \tag{8.63}$$

where

$$Q = \frac{R_P}{X_P} = \frac{X_S}{R_S} \tag{8.64}$$

Using the Norton equivalent of the inductor, its resistance and the voltage induced by the reader gives the equivalent circuit shown in Figure 8.69.

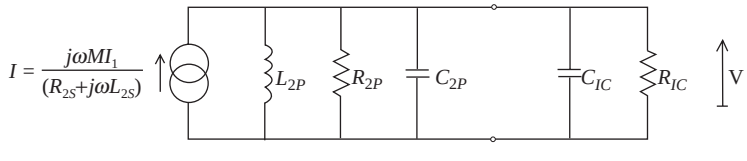


Figure 8.69 Parallel equivalent circuit of a tag antenna and IC

With an equivalent circuit consisting of only shunt-connected components, the resonant frequency is simple to calculate. It is given by

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{L_{2P}C_T}} \tag{8.65}$$

where C_T is the combined capacitance

$$C_T = C_{IC} + C_{2P} \tag{8.66}$$

It is also clear that maximum power transfer occurs at the resonant frequency when R_{2P} is equal to R_{IC}. The voltage developed across the load resistor is given by

$$|V_L| = \frac{\sqrt{1+Q^2}}{Q} \frac{\omega M\,|I_1|}{\sqrt{\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]^2 + \left[\frac{\omega L_{2P}}{R_T}\right]^2}} \tag{8.67}$$

where R_T is the combined resistance of R_{2P} in parallel with R_{IC},

$$R_T = \frac{R_{IC}R_{2P}}{R_{IC}+R_{2P}} \tag{8.68}$$

Again this is a maximum at the resonant frequency given by Equation (8.65). At resonance, and for a given reader current and tag-to-reader separation, the voltage is predominantly dependent on R_T, which, in turn, is usually dominated by the resistance of the tag coil, R_{2P}. This means that the read range (or the range over which there is a voltage great enough to be rectified) is determined by the tag coil quality.

LF and HF tag ICs are designed to have a capacitive input impedance that can be made to resonate with a typical tag coil. Note that tag ICs are normally complementary metal oxide semiconductor (CMOS) devices, which are inherently capacitive. When resonant, the tags produce a high voltage that is then rectified in order to provide a DC power supply to the rest of the tag IC. Once powered, the tag is able to switch its input impedance in order to produce an impedance change at the reader coil. Normally, switching is between the inherent capacitive reactance of the tag IC and a short circuit – an open circuit is seldom used in passive tags due to the high input voltage that could be generated. The impedance change that occurs at the terminals of the reader antenna can be detected by measuring the change in voltage or current that occurs. Hence, digital information can be transferred from the tag to the reader.

A typical rectangular tag antenna is shown in Figure 8.70.

The inductance of coils of the type shown in Figure 8.70 can be approximated by the formulas that follow [27].

$$L = \frac{\mu_0}{\pi}\left[x_1 + x_2 + x_3 + x_4\right] N^p \tag{8.69}$$

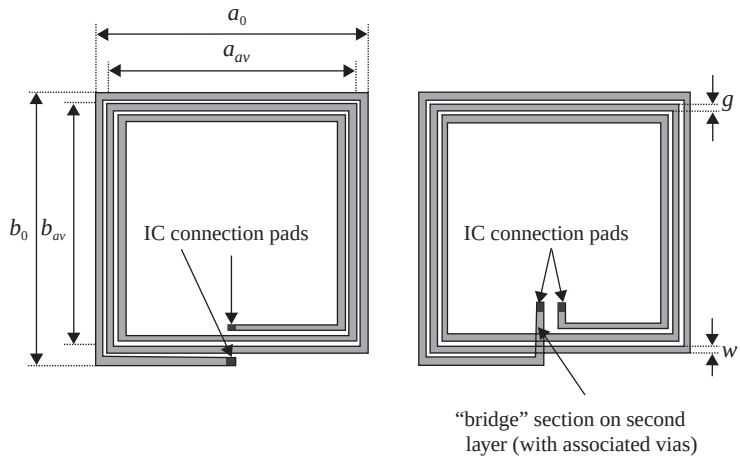


Figure 8.70 Single- and two-layer tag coils

where

$$x_1 = a_{av} \ln \left[\frac{2a_{av}b_{av}}{d\left(a_{av} + \sqrt{a_{av}^2 + b_{av}^2}\right)} \right] \tag{8.70}$$

$$x_2 = b_{av} \ln \left[\frac{2a_{av}b_{av}}{d\left(b_{av} + \sqrt{a_{av}^2 + b_{av}^2}\right)} \right] \tag{8.71}$$

$$x_3 = 2\left[a_{av} + b_{av} - \sqrt{a_{av}^2 + b_{av}^2} \right] \tag{8.72}$$

$$x_4 = \frac{a_{av} + b_{av}}{4} \tag{8.73}$$

$$d = \frac{\mathbf{2}\,(t + w)}{\pi} \tag{8.74}$$

Here a_0, a_{av}, b_0, b_{av}, g and w are as shown in Figure 8.70. The track thickness is denoted by t, N is the number of turns and p is the turn exponent, which is dependent on the manufacturing technology, as indicated in Table 8.8.

Table 8.8 Turn exponent with antenna technology

Coil technology	Turn exponent, p
Wires	1.8–1.9
Etched tracks	1.75–1.85
Printed tracks	1.7–1.8

These formulas can be used as a starting point for a design. Any final design is likely to be simulated using a 2.5D or 3D electromagnetic simulator, so that the effects of finite conductivity, dielectric materials, etc. can be accounted for.

An example of a commercial coil antenna – used in a European passport – is shown in Figure 8.71. The antenna has five turns formed using thin wires. The chip is connected to a metallic strap prior to connection to the coil. The IC is also embedded in a relatively thick layer of insulating material. This prevents the IC from being damaged and also prevents the coil from being excessively bent.

8.4.3 Far-Field Systems

Systems with carrier frequencies above 100 MHz generally operate by transferring power in the far field. The most commonly used frequency bands are in the region of 900 MHz, though the ISM band centered at 2.45 GHz is also used (this band is sometimes termed microwave rather than UHF). Some typical frequency bands for different parts of the world are indicated in Table 8.9.

Note that for worldwide operation in the region of 900 MHz, a wide frequency range – approximately 860 to 960 MHz – is required. This means that the antenna must be relatively wideband, with a fractional bandwidth of approximately 11%. Because of this, the antenna Q

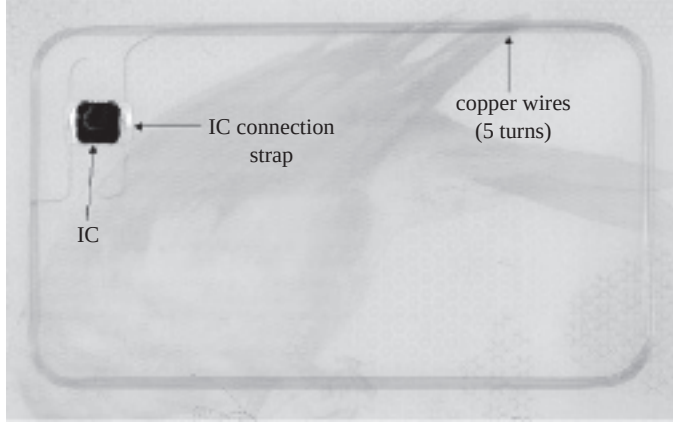


Figure 8.71 Coil antenna used within a passport

Table 8.9 RFID frequencies commonly used within the UHF band

Region	Frequencies (MHz)	Power	Comment
Europe	865.6–867.6	2 W ERP	License required in some countries
North America	902–928	4 W EIRP	Applies to the United States and Canada
South America	902–928	4 W EIRP	
Africa	865.6–867.6	2 W ERP	
India	865–867	4 W ERP	
China	917–922	2 W ERP	Provisional. Temporary license required.
Japan	952–954	4 W EIRP	License required.
Worldwide	2400–2483.5		These frequencies apply to the USA and most of Europe. Slightly different frequencies are used in some countries

is limited, leading to a limited voltage at the tag IC. Importantly, this indicates that there is a trade-off between bandwidth of operation and range.

A typical far-field system is shown in Figure 8.72. In contrast to near-field systems, coil antennas are rarely used in far-field systems. Indeed, a wide variety of antennas is possible. Dipoles, PIFAs and patches are amongst the options at the reader. Patches are often used to provide circular polarization (as illustrated in Figure 8.72): this helps to make the communication with an arbitrarily oriented, linearly polarized tag antenna less variable. Tag antennas are rarely circularly polarized (the cost of fabrication is prohibitively high); modified dipoles are most commonly used.

The transfer of energy between far-field antennas can be evaluated by consideration of Figure 8.73.

Use Friis's transmission formula, discussed in Chapter 4 and [28], to give

$$\frac{P_R}{P_T} = \left(\frac{\lambda}{4\pi\, d}\right)^2 G_T\, G_R \tag{8.75}$$

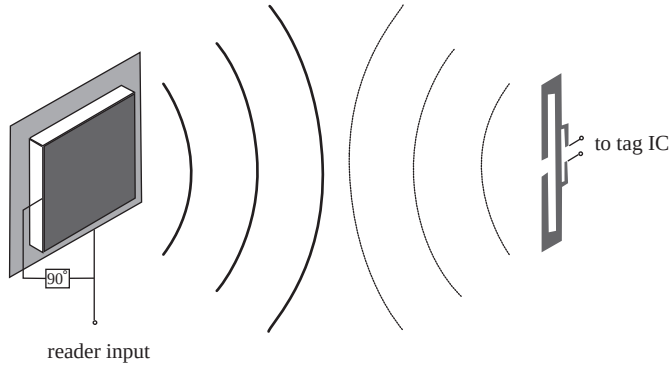


Figure 8.72 Typical far-field reader and tag antennas

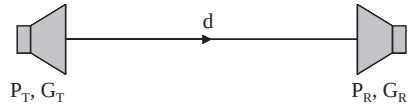


Figure 8.73 Pathloss model

This formula applies to perfectly matched antennas with the same polarization. Correction factors, $|\tau|^2$ and χ must be applied to compensate for the effects of mismatch and polarization misalignment respectively. Including these effects, the power received by the tag antenna is given by

$$P_R = \left(\frac{\lambda}{4\pi\, d}\right)^2 P_T G_T G_R |\tau|^2 \chi \tag{8.76}$$

For the tag to operate effectively, it must receive enough power from the tag to 'power up' and it must be capable of providing sufficient modulation for the reader to detect. Both depend on the power of the reader, the antenna gains, the polarization matching between the reader and tag antennas and the mismatch between the tag antenna and the associated IC. The former also depends on the power required to energize the tag, P_{CHIP}, whereas the latter also depends on the effectiveness of the load switching and the sensitivity of the reader. The reader sensitivity is easier to improve than the power required to energize the tag, since the reader is a more elaborate and expensive device. Hence, the range, d, of the system is normally determined by the minimum power required by the tag IC, given by

$$d = \frac{\lambda}{4\pi}\sqrt{\frac{P_T G_T G_R |\tau|^2 \chi}{P_{CHIP}}} \tag{8.77}$$

To improve range, tag ICs are designed to have the lowest possible P_{CHIP}. The range of a commercial device is calculated in the following example. A typical tag IC has an equivalent series input resistance of $10 - j245$ ohms. The tag antenna is designed such that its impedance is as close as possible to the complex conjugate of the chip input impedance (or the impedance at which the tag rectification efficiency is greatest). Hence, an antenna impedance of $10 + j245$ ohms is required. The resistance is low, which conveniently suits electrically small antennas such as dipoles operated well below half a wavelength. However, the required inductive reactance does not suit small dipoles: such antennas are normally capacitive. Fortunately, it is straightforward to modify the impedance of a dipole in order to make it inductive. A simple technique is to use a shunt feed, as illustrated in Figure 8.74. The length of the dipole, l_1, is less than $\lambda/2$, giving a low resistance and a capacitive reactance. The shunt arms provide both an impedance transformation (increasing both the resistance and the capacitive reactance) and a shunt inductance, the combination of which can be used to transform the antenna impedance to the desired value. The impedance transformation is predominantly determined by the ratio of w_2 to w_1 (the higher this ratio, the greater the transformation). The shunt inductance is predominantly determined by the length l_2(though w_1, w_2 and s_1 also have some influence).

Variations of the shunt-fed dipole, shown in Figure 8.74, are often used in practice.

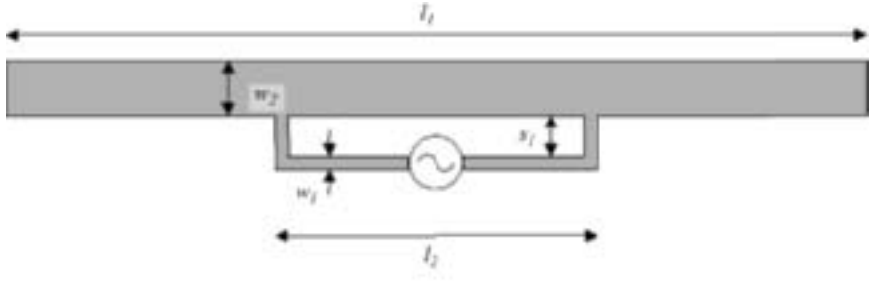


Figure 8.74 Variations of the shunt-fed dipole

8.5 Reconfigurable Antennas

8.5.1 Introduction

Work has been performed on reconfigurable antennas for many years [29]. As the term suggests, reconfigurable antennas employ some form of variability – often by using switches, variable components or moving parts – to change the resonant frequency, radiation pattern and/or impedance match.

For mobile radio, moves towards software-defined radios (SDRs) are driving research on reconfigurable antennas [30]. The Federal Communications Commission (FCC) – an independent United States government agency that regulates radio communications – definition of an SDR is as follows [31]:

> 'a radio that includes a transmitter in which the operating parameters of the transmitter, including the frequency range, modulation type or maximum radiated or conducted output power can be altered by making a change in software without making any hardware changes.'

There are many ways of reconfiguring an antenna, including:

- resonant mode switching/tuning (via shorting, reactive loading);
- feed network switching/tuning;
- mechanical reconfiguration (with moving parts!)

Each of these methods is considered in the sections that follow. However, it is useful to first consider what switching and variable components are available.

8.5.2 Switching and Variable-Component Technologies

Figure 8.75 shows a simplified equivalent circuit of a switch. In the 'ON' state, the switch can be represented by a simple resistor, whereas in the 'OFF' state, it can be characterized by a series capacitor. For many switches, this capacitor is not lossless, so a parallel resistance is included to indicate that the capacitor has a finite quality factor, Q.

The insertion loss of single-pole, single-throw and series-connected, single-pole, dual-throw (SPDT) switches is determined primarily by the 'ON' resistance, R_{ON}. The isolation is predominantly determined by the 'OFF' capacitance, C_{OFF}. Generally, there is a trade-off between

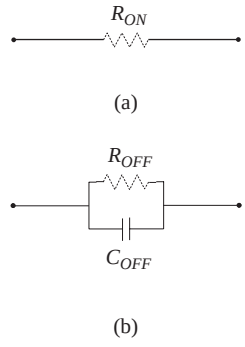


Figure 8.75 Simple switch equivalent circuit (a) in the 'ON' state; (b) in the 'OFF' state

R_{ON} and C_{OFF} such that their product, $R_{ON}C_{OFF}$ is constant for a particular technology and DC power supply. Hence, the isolation is usually improved at the expense of insertion loss and vice versa.

Table 8.10 gives a comparison of some switch technologies: PIN diodes, Gallium Arsenide (GaAs) and Indium Phosphide (InP) field effect transistors (FETS) and micro electromechanical systems (MEMS).

PIN diodes have been used with good results in many applications. They have the advantage of potentially low 'ON' resistance, but this comes at the cost of high current consumption, which is a disadvantage for low-power applications – in particular, for portable devices such as mobile phones. PIN diodes can have a reasonable isolation, but this requires a large negative bias. Again, this is inconvenient in portable devices, where voltages of between around 0 and 3 V are available.

GaAs and InP FETs have become increasingly popular in portable devices such as mobile phones because they offer low current consumption. Though the $R_{ON}C_{OFF}$ product of GaAs and InP FETs is approximately the same as for PIN diodes, FETs cannot handle large voltages and are, hence, often used in series. This increases R_{ON} and reduces C_{OFF} and often means that switches made with GaAs and InP FETs have higher 'ON' resistances and lower 'OFF' capacitances than their PIN-based equivalents.

MEMS devices are in their commercial infancy. They can be likened to miniature relays where metal-to-metal contacts are made and broken by the application (or absence) of an electrostatic field (in contrast to reed relays, where actuation is via a magnetic field). Unlike

Table 8.10 Comparison of switch technologies

Device	'ON' resistance	Current consumption	Isolation	'OFF' resistance	Linearity
PIN diodes	low	low	reasonable	appreciable	reasonable
GaAs and InP FETs	reasonable	low (approx. 100 μA)	good	reasonable	good
MEMS	potentially very low	very low (a few μA)	very good	high (high Q)	very good

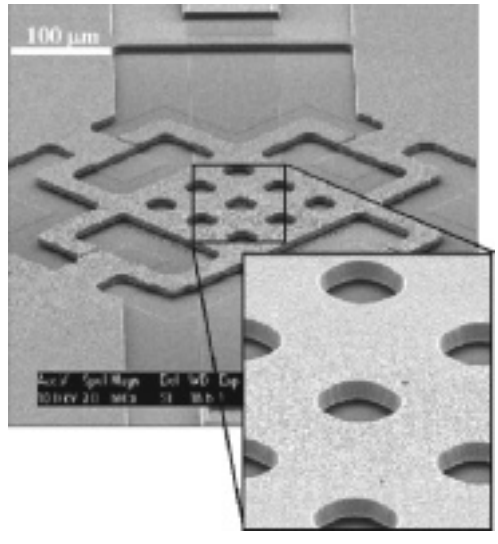


Figure 8.76 Fabrication detail of a typical MEMS switch

conventional relays, MEMS devices are very small and are fabricated (as far as is possible) using processes that are normally used to manufacture low-cost integrated circuits. Figure 8.76 shows an example. This manufacturing technique, coupled with near lossless 'ON' and 'OFF' states – with metal-to-metal contacts and an air gap respectively – makes MEMS-based switches highly attractive for future applications. The main drawback of MEMS devices is that they typically require high DC actuation voltages.

Switches are not the only components that can be used to make reconfigurable antennas. If large components can be tolerated, mechanically variable inductors and capacitors can be utilized. For example, variable capacitors – where rotary motors provide the variation – are widely used within high-power antenna-tuning units in the MF and HF bands. On a smaller scale, varactor diodes may be used, though they generally suffer from poor linearity and are therefore limited to low-power applications. Active devices may also be used, for example as variable inductors or capacitors. The main drawbacks of doing so are that the active devices consume power, add noise and are often nonlinear. In the future, MEMs devices may also be used as variable capacitors, though they tend to have rather nonlinear capacitance curves with DC voltage. MEMS devices may also be used, not as galvanic switches (where a metal-to-metal contact is made), but as capacitive switches with a high 'ON' to 'OFF' capacitance ratio. This avoids problems associated with the 'sticking' and wear of very small metallic contacts.

8.5.3 Resonant Mode Switching/Tuning

Figure 8.77 shows an example of resonant mode tuning, where a PIFA antenna is tuned by a variable capacitor. As the capacitance value is increased, the antenna resonant frequency is reduced (with the penalty of reduced bandwidth). The capacitor is located at the point on the

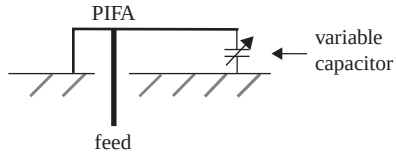


Figure 8.77 A side view of a PIFA antenna that is tuned by a variable capacitance

antenna where the electric field is highest, which gives the highest possible resonant frequency tuning ratio for a given capacitance variation.

For physically large (low-frequency) applications, the capacitor could be a mechanically variable capacitor; for low-power applications or applications where high degrees of linearity are not required, it could be a varactor diode. Alternatively, discrete variability could be provided by switched capacitors.

8.5.4 Feed Network Switching/Tuning

Antennas can be reconfigured either on the antenna itself (as in the previous section) or by modifying the circuitry feeding the antenna. An obvious example of the latter is the phased array. Here, the phases and amplitudes of the signals applied to each element of the array can be modified. In turn, this can alter the antenna radiation pattern in terms of its angle of maximum gain, the value of the gain and the pattern shape.

8.5.5 Mechanical Reconfiguration

Figure 8.78 shows an example of mechanical reconfiguration [32]. A 'V'-shaped dipole is implemented and a MEMS micro-hinge is used to vary the angle of the V. This allows the antenna to beam-steer.

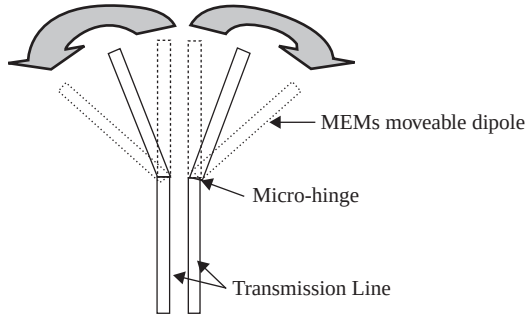


Figure 8.78 A 'V' dipole with MEMS-enabled beam-steering

8.6 Summary

In this chapter we have discussed five special topics: electrically small antennas, mobile antennas, multiband and UWB antennas, RFID antennas and reconfigurable antennas. An in-depth study of electrically small antennas was conducted. The focus was on impedance bandwidth theory and broadening techniques. Antenna diversity and user interaction were also dealt with, both in theory and by studying practical examples. As new wireless systems are emerging all the time, there are always demands for smarter and/or smaller antennas.

References

[1] H. Wheeler, 'Small Antennas,' *IEEE Trans. Antennas and Propagation*, **AP-23**(4), 462–469, 1975.
[2] G. L. Matthaei, L. Young, and E.M.T. Jones, *Microwave Filters, Impedance Matching Networks and Coupling Structures*, McGraw-Hill, 1964.
[3] H. Wheeler, 'Fundamental Limitations of Small Antennas,' *Proc. IRE*, December 1947, pp. 1479–1484.
[4] L. J. Chu, 'Physical Limitations of Omni-Directional Antennas,' *J. Appl. Phys*., **19**, 1163–1175, 1948.
[5] R. Harrington, 'Effects of Antenna Size on Gain, Bandwidth, and Efficiency,' *J. Nat. Bur. Stand.*, **64-D**, 1–12, 1960.
[6] R. E. Collin and S. Rothchild, 'Evaluation of Antenna *Q*,' *IEEE Trans. AP*, **AP-12**, 23–27, 1964.
[7] R. L. Fante, 'Quality Factor of General Ideal Antennas,' *IEEE Trans. AP*, **AP-17**, 151–157, 1969.
[8] J. S. McLean, 'The radiative properties of electrically-small antennas,' *IEEE Int. Sympos. on Electromagnetic Compatibility*, 1994, pp. 320–324.
[9] J. S. McLean, 'A re-examination of the fundamental limits on the radiation *Q* of electrically small antennas,' *IEEE Trans. AP*, **44**(5), 672–676, 1996.
[10] H. D. Foltz and J. S. McLean, 'Limits on the radiation *Q* of electrically small antennas restricted to oblong bounding regions,' *Antennas and Propagation Society International Symposium 1999*, **4**, 2702–2705.
[11] G. A. Thiele, P. L. Detweiler and R. P. Penno, 'On the lower bound of the radiation *Q* for electrically small antennas,' *IEEE Trans. AP*, **51**(6), 1263–1269, 2003.
[12] H. W. Bode, 'Network Analysis and Feedback Amplifier Design,' Van Nostrand, 1947.
[13] R. M. Fano, 'Theoretical limitations on the broadband matching of arbitrary impedances,' *J. Franklin Institute*, **249**(1), 57–83, 1950.
[14] R. M. Fano, 'Theoretical limitations on the broadband matching of arbitrary impedances,' *J. Franklin Institute*, **249**(2), 139–154, 1950.
[15] G. Goubau, 'Multi-element Monopole Antennas,' *Proc. ECOM-ARO Workshop on Electrically Small Antennas*, Fort Monmouth, May 1976, pp. 6–7.
[16] J. D. Parsons, *The Mobile Radio Propagation Channel*, 2nd edition, John Wiley & Sons, Ltd, 2000.
[17] R. King, C. W. Harrison and D. H. Denton, 'Transmission line missile antennas,' *IRE Trans. Antennas and Propagation*, **8**, 88–90, 1960.
[18] H. Mishima and T. Taga, 'Mobile antennas and duplexer for 800 MHz band mobile telephone system,' *Antennas and Propagation Society International Symposium 1980*, **18**, 508–511.
[19] K. Ogawa and T. Uwano, 'A diversity antenna for very small 800-MHz band portable telephones,' *IEEE Trans. AP*, **42**(9), 1342–1345, 1994.
[20] P. Massey, 'Finite element simulation of SAR,' *IEE Antenna Measurement and SAR Seminar (IEE AMS 2002)*, Loughborough University, 28–29 May, 2002.
[21] A. Hottinen, O. Tirkkonen and R. Wichman, *Multi-antenna Transceiver Techniques for 3G and Beyond*, John Wiley & Sons, Ltd, 2003.
[22] G. Tsoulos (Ed.) *MIMO System Technology for Wireless Communications*, CRC, 2006.
[23] R. Aiello and A. Batra, *Ultra Wideband Systems: Technologies and Applications*, Elsevier Inc., 2006.
[24] B. Allen, M. Dohler, E. Okon, W. Malik, A. Brown and D. Edwards (Eds), *Ultra Wideband Antennas and Propagation for Communications, Radar and Imaging*, John Wiley & Sons, Ltd, 2006.

[25] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*, 2nd edition, John Wiley & Sons, Ltd, 2003.
[26] S. Ramo, J. Whinnery and T. Van Duzer, *Fields and Waves in Communication Electronics*, 3rd edition, John Wiley & Sons, Inc., 1993.
[27] 'I•CODE Coil Design Guide,' Philips Semiconductors Application Note, 2002.
[28] H. T. Friis, 'A note on a simple transmission formula,' *Proc. IRE,* **41**, 254–256, 1946.
[29] J. T, Bernhard, *Reconfigurable Antennas*, Morgan & Claypool, 2006.
[30] J. H. Reed, *Software Radio: A Modern Approach to Radio Engineering*, Prentice Hall, 2002.
[31] FCC's Notice of Proposed Rulemaking, ET Docket 00-47, issued December 8, 2000.
[32] J.-C. Chiao, Y. Fu, I. M. Chio, M. DeLisio and L.-Y. Lin, 'MEMS reconfigurable Vee antenna,' *IEEE Microwave Symposium Digest*, **4**, 1515–1518, 1999.

Index

A
Absorbing materials: see Radio absorbing materials
Admittance, 24
Adaptive array: see Antenna arrays
Ampere's circuital law, 18
Anechoic chambers, 267
Antenna analysis methods, 217
Antenna aperture, 118
Antenna arrays, 191–203
  broadside array, 196
  element coupling, 200
  end-fire array, 197
  isotropic linear array, 192
  pattern multiplication principle, 199
  uniform array, 193
Antenna definition 14
Antenna diversity, 318–324
  combing methods, 320
  correlation coefficient, 319
  polarization diversity, 319
  radiation pattern diversity, 319
  spatial diversity, 319
Antenna factor (AF), 120
Antenna history, 1–4
Antenna impedance: see Antenna input impedance
Antenna input impedance, 122
  measurements, 261
Antenna measurements, 256–281
  anechoic chamber, 267
  calibration, 260
  compact antenna test range (CATR), 268
  efficiency, 273
  free space ranges, 265
  gain, 271
  impedance; 261
impedance de-embedding techniques, 275
  near field methods, 270
  probe array near-field system, 276
  open area test site (OATS), 265
  radiation pattern, 268–272
  reverberation chamber, 274
Antenna pattern: see Radiation patterns
Antenna polarization, 122, 206
  circular polarisation, 207
Antenna radiation resistance, 123
Antenna ranges: see Antenna measurements
Antenna size reduction, 299
  dielectric loading, 303
  matching, 301
  reactive loading, 302
  top loading, 299
Antenna system, 15
Antenna temperature, 120
Antenna user interaction, 325–333
Aperture efficiency, 118
Aperture-type antennas, 163–191
Arrays: see Antenna arrays
Array factor, 193
Artificial materials, 256
Attenuation constant, 28, 78, 95, 102
AWS, 305
Axial ratio, 83, 148

© 2008 John Wiley & Sons, Ltd

B
Babinet's principle, 183
Baluns, 205, 249
Bandwidth, 51–55, 125, 283–298
  broadening, 289
  fractional bandwidth, 52
  limits, 289
Basis functions, 218
Bionical antenna, 137
Biot-Savart law, 15
Bit error rate (BER), 101
Bode-Fano limit, 51
Bow-tie antenna, 137
Boundary conditions, 19
Brewster's angle, 87
Broadside arrays, 196

C
Capacitance, 24
Channels, 100
  Gaussian, 100
  Rician, 101
  Rayleigh, 101
Characteristic impedance, 28
Circuit concepts, 23
Coaxial cable, 57–60
  cut-off frequency, 59
Coherence bandwidth, 100
Complex conjugate, 44
Complex numbers, 6–7
Complex permittivity: see Permittivity
Computer aided design, 215, 233–250
Conductance, 24
Conductivity, 13, 313
  conductivity table, 14
Connectors, 70–74
Convergence of numerical results, 227
Coordinates, 10
  Cartesian coordinates, 10
  spherical coordinates, 10
Coplanar waveguide (CPW), 66–68
Critical angle, 87
Cumulative distribution function (CDF), 320
Current element radiation, 108–112

D
dB, 12
dBi, 116
dBd, 116
DCS, 305
Delay spread, 100
Dielectric constant: see Permittivity
Dielectric resonant antennas (DRA), 4, 255
Diffraction, 91
Dipoles, 129–137
  current distribution, 130
  directivity, 132
  half-wavelength, 135
  impedance, 133
  radiation efficiency, 136
  radiation pattern, 131
  short dipole, 135
  summary, 131
Directivity, 115
Distributed element system, 24
Diversity: see Antenna diversity
Diversity gain, 316, 321
  mean effective gain (MEG), 316
  also see Antenna diversity
Dolph-Tchebyscheff (D–T) optimum distribution, 198
Double-ridged horn, 163, 175
Digital signal processing (DSP), 325
Duality principle, 141

E
Effective aperture, 118
Effective height, 119
Effective isotropic radiated power (EIRP), 117
Effective permittivity, 67
Effective radiated power, 118
Effects of ground plane, 139
Electric displacement: see Electric flux density
Electric current, 23
Electric field, 12
Electric flux density, 13
Electric force, 13, 15
Electrically small antennas, 108–112, 283–303

Electromagnetic compatibility (EMC), 110
EM spectrum, 11
  cellular radio frequencies, 306
  ISM band, 11
  UHF, 11
  VHF, 11
End-fire arrays, 196
  bandwidth, 198
Equivalence principle, 92
EZNEC, 234

F
Fading; 100
  flat fading, 100
  frequency selective, 100
Far field, 109–113
Far field condition, 111, 269
Faraday's law of induction, 17
Faraday rotation, 96
Field concepts, 77
Finite-difference time-domain method (FDTD), 229
Finite element methods, 228
First Fresnel zone distance, 99
Flip antenna, 308, 315, 327
Folded dipole, 137
Fourier transforms, 163
Fraunhofer region: see Far field
Frequency domain methods, 231
Frequency independent antennas, 161
Fresnel region, 112
Friis' transmission formula, 97, 120

G
Gain, 116
  field gain, 116
  power gain, 116
Galerkin's method, 220
Galvanic corrosion, 254
Gauss' Law, 18
Geometrical optics (GO), 93, 232
Geometrical theory of diffraction (GTD), 232
Global positioning system (GPS), 303, 306
Global systems for mobile (GSM), 244, 249, 305
Ground plane: see Effects of ground plane

H
Half-power beamwidth (BPBW), 114
Hansen-Woodyard condition, 149
Hansen-Woodyard end-fire array, 197
Helical antennas, 140, 147–152, 309
  axial mode, 148
  normal mode, 147
Hertz, Heinrich, 1
High frequency methods, 217, 232
High frequency structure simulator (HFSS), 244
HIPERLAN, 306
Horn antennas, 169,
  also see Pyramidal horns
Human body effects, 325
Huygens' principle, 91

I
Image theory, 137
Impedance, 23
  bandwidth, 283
  measurements, 261
  mutual impedance, 201
  self impedance, 201
Impedance matching, 44–51, 283–298
  distributed matching network, 44–47
  double tuning, 284, 289
  lumped matching network, 47–51
Increased directivity end-fire array, 197
Inductance, 23
Inner product, 219
Input impedance, 31, 122
Inter-symbol interference (ISI), 100
Intrinsic impedance, 81, 111
Inverted F antenna, 140, 311
  also see PIFA
Inverted L antenna, 140, 310
Ionosphere, 95
Isotropic antenna, 113

L
Lens antennas, 180
Liquid crystal polymer (LCP), 255

Logarithmic scales, 12
Log-periodic antennas, 157
  apex angle, 158
  design, 158
  operation principle, 157
  scaling factor, 158
  spacing factor, 158
Log-periodic dipole antenna (LPDA), 157
Loops, 141–147
  directivity, 144
  impedance, 143, 146
  one-wavelength loop, 144
  radiation pattern, 143, 145, 146
  small loops, 142
Lorentz force, 16
Loss tangent, 13
low-temperature co-fired ceramics (LTCC), 64
Lumped element system, 24

M
Magnetic field, 15
Magnetic flux density, 15
Marconi, Guglielmo, 1, 304
Matching efficiency, 116, 124, 326
Maxwell, James Clerk, 16
Maxwell's equations, 16–19
Mean effective gain (MEG), 316–333
Meander line antennas, 302, 305
Method of moment (MoM), 218–228
Micro electromechanical
systems (MEMS), 352
Microstrip antennas, 184–191
  bandwidth, 188
  design, 188
  directivity, 187
  ground plane, 191
  input impedance, 187
  radiation pattern, 181
Microstrip line, 60–63
  cut-off frequency, 63
Mie theory, 93
Mismatch efficiency, 326
Mobile antennas, 305–318
Mobile radio, 304, 306
Modulated scattering techniques, 271
Monolithic microwave integrated circuit, (MMIC), 184
Monopole antennas, 137–141, 307, 308
  ground effects, 139
  summary, 139
Multi-band antennas, 333–336
  higher order resonances, 333
  resonant traps, 334
  combined resonant structures, 334
  parasitic resonators, 335
Multi-path fading, 100, 315
Mutiple-in multiple-out (MIMO) antennas, 4, 101, 274, 324
Mutual coupling, 200

N
Near field, 111–113
Network analyzer, 258
Numerical electromagnetic code (NEC), 216
Numerical methods, 217–231

O
Ohm's law, 24
Open area test site (OATS), 265
Open-ended waveguide, 166–168

P
Paraboloidal reflector antennas, 175–180
  analysis and design, 176
  aperture efficiency, 177
  directivity, 177
  off-set parabolic reflectors, 179
Path loss, 97
  free space, 97
  Two-ray model, 98
  multi-path model, 99
Pattern: see Radiation patterns
Pattern multiplication principle, 199
PCS, 305
Permeability, 15
  relative permeability table, 16
Permittivity, 13, 19, 60, 67, 313
  relative permittivity table, 14
  effective relative permittivity, 60, 67
Personal area network (PAN), 306

Phantom, 313
Phase constant, 28, 29, 78
Phased arrays, 194
Physical optics (PO), 232
Physical theory of diffraction (PTD), 233
Planar inverted F antennas (PIFA), 244, 307, 310
Plane wave, 80
Pocklington's integral equation, 224
Polarization, 82, 122
  circular polarization, 82
    left-hand circular polarization, 83
    right-hand circular polarization, 82
  linear polarization, 82
  parallel polarization, 84
  perpendicular polarization, 84
Power, 24
Power conservation law, 257
Power density, 81
Power density function (PDF), 100
Poynting vector, 81
Printered circuit board (PCB), 307, 312
Probability density function (PDF), 317
Propagation
  in rain, 96
  in fog, 97
  in snow, 97
  models, 97
  path-loss: see Path loss
Propagation channels: see Channels
Propagation constant, 27, 28, 78
Pyramidal horns, 169–175
  E-plane sectorial horn, 169
  H-plane sectorial horn, 169
  optimum design, 169, 171

Q
Q factor: see Quality factor
Quality factor, 51–55, 284–298
Quarter-wavelength transform, 34, 36
Quasi-TEM mode, 62, 68

R
Radar cross section (RCS), 121
Radiating near field, 111
Radiation efficiency factor, 117, 124
Radiation intensity, 115
Radiation patterns, 112–115
  directivity: see Directivity
  E-plane, 113
  field pattern, 113
  first null beamwidth, 114
  first side lobe level, 114
  front-to-back ratio, 114
  gain: see Gain
  H-plane, 113
  half-power beamwidth (HPBW), 114
  power pattern, 113
Radiation resistance, 123
Radio absorbing materials (RAM), 261
Radio channels: see Channels
Radio communication system, 5–6
Radio frequency identification (RFID), 110, 256, 339–350
  far-field system, 348
  near-field system, 342
  reader, 340
  tags, 340
Radome, 208
Ray optics, 232
Rayleigh scattering, 93
Reactance, 23
Reactive near field, 111–113
Reciprocity theorem, 203
Reconfigurable antennas, 351–354
Reflection: see Wave reflection
Reflection coefficient, 31, 36, 123
  measurements, 261
Reflector antennas, 175
Relative dielectric constant: see Permittivity
Relative permeability: see Permeability
Relative permittivity: see Permittivity
Refraction, 83
Resistance, 23
Resistivity, 13
Return loss, 36, 124
  measurements, 261

S
Scattering, 92
Scattering parameters, 256
Self-complimentary antennas, 162

Signal to noise ratio (SNR), 101, 205
Skin depth, 101–104, 327
Sleeve dipole, 137
Slope parameters, 283
Slot antennas, 180
Smart antennas, 194
Smith Chart, 41
Snell's law, 84
Software defined radio (SDR), 351
Speed of light, 11, 79, 80
Spherical coordinates, 20
Specific absorption rate (SAR), 307, 310, 312
Spectrum: see EM spectrum
Stripline, 64
Substrate, 64
  table, 64
Surface resistivity, 64
Surface wave/mode, 63
Susceptance, 24

T
TDR, 260
TEM horn, 163
TEM mode, 57, 59, 66
TE mode, 68
Telegraph equation, 27
Testing functions, 219
TM mode, 68
Time domain methods, 231
Top-loaded antenna, 140
Transmission line modelling (TLM) method, 230
Transmission lines, 25–70
  characteristic impedance, 28
  input impedance 31
  lossless, 30
  low loss, 31
  model, 25–28
  quarter-wavelength transform, 34, 36
  reflection coefficient, 31
  transmission line equation, 27
  velocity, 30
Two-wire transmission line, 55–57
  twisted-pair transmission line, 57

U
Ultra-wideband (UWB) antennas: see Wideband antennas
UMTS, 305
Uniform theory of diffraction (UTD), 232

V
Vector network analyzer (VNA), 260
Vector operation, 7–10
  commutative law, 8
  cross product, 9
  curl operator, 17
  divergence operator, 17
  dot product, 8
  right-hand rule 9
  scalar product, 8
  vector product, 9
Vectors, 7
Voltage, 23
Voltage standing wave ratio (VSWR), 38, 124
  measurements, 261

W
Wave equation, 78
Wave length, 11, 34
Wave number: see Phase constant
Wave optics, 232
Wave reflection, 83
  reflection coefficient, 83
  reflection on a conductor, 85
  reflection on a ground, 85
  reflection of a wall, 89
Wave transmission, 83
  transmission coefficient, 84
Wave velocity, 11
Waveguide, 68–70
  cut-off frequency, 69
  standard waveguides, 70
  TE_{mn} mode, 68
  TM_{mn} mode, 68
Weighting functions, 219
Wheeler cap, 274
Wideband antennas, 73, 336–339
WiMAX, 3–4

Wireless local area network (WLAN), 306
Wire-type antennas, 129–163

Y
Yagi-Uda antenna, 152
  current distribution, 154
  design, 156
  directivity, 155
  director, 153
  driven element, 153
  input impedance, 155
  operation principle, 152
  radiation pattern, 154
  reflector, 153